

Northumbria Research Link

Citation: Greenhalgh, Joanne, Dalkin, Sonia, Gooding, Kate, Gibbons, Elizabeth, Wright, Judy, Meads, David, Black, Nick, Valderas, Jose Maria and Pawson, Ray (2017) Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Health Services and Delivery Research, 5 (2). pp. 1-280. ISSN 10.3310/hsdr05020

Published by: NIHR Health Technology Assessment Programme

URL: <http://dx.doi.org/10.3310/hsdr05020> <<http://dx.doi.org/10.3310/hsdr05020>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/29544/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care

Joanne Greenhalgh, Sonia Dalkin, Kate Gooding, Elizabeth Gibbons, Judy Wright, David Meads, Nick Black, Jose Maria Valderas and Ray Pawson



***National Institute for
Health Research***

Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care

Joanne Greenhalgh,^{1*} Sonia Dalkin,² Kate Gooding,¹ Elizabeth Gibbons,³ Judy Wright,¹ David Meads,¹ Nick Black,⁴ Jose Maria Valderas⁵ and Ray Pawson¹

¹School of Sociology and Social Policy, University of Leeds, Leeds, UK

²Department of Public Health, Northumbria University, Newcastle upon Tyne, UK

³Nuffield Department of Population Health, University of Oxford, Oxford, UK

⁴Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

⁵Institute of Health Research, University of Exeter, Exeter, UK

*Corresponding author

Declared competing interests of authors: none

Disclaimer: This report contains transcripts of interviews from studies identified during the course of the research and contains language that may offend some readers.

Published January 2017

DOI: 10.3310/hsdr05020

This report should be referenced as follows:

Greenhalgh J, Dalkin S, Gooding K, Gibbons E, Wright J, Meads D, *et al*. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. *Health Serv Deliv Res* 2017;**5**(2).

Health Services and Delivery Research

ISSN 2050-4349 (Print)

ISSN 2050-4357 (Online)

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HS&DR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hsdr. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Services and Delivery Research* journal

Reports are published in *Health Services and Delivery Research* (HS&DR) if (1) they have resulted from work for the HS&DR programme or programmes which preceded the HS&DR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

HS&DR programme

The Health Services and Delivery Research (HS&DR) programme, part of the National Institute for Health Research (NIHR), was established to fund a broad range of research. It combines the strengths and contributions of two previous NIHR research programmes: the Health Services Research (HSR) programme and the Service Delivery and Organisation (SDO) programme, which were merged in January 2012.

The HS&DR programme aims to produce rigorous and relevant evidence on the quality, access and organisation of health services including costs and outcomes, as well as research on implementation. The programme will enhance the strategic focus on research that matters to the NHS and is keen to support ambitious evaluative research to improve health services.

For more information about the HS&DR programme please visit the website: <http://www.nets.nihr.ac.uk/programmes/hsdr>

This report

The research reported in this issue of the journal was funded by the HS&DR programme or one of its preceding programmes as project number 12/136/31. The contractual start date was in March 2014. The final report began editorial review in December 2015 and was accepted for publication in May 2016. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HS&DR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HS&DR programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2017. This work was produced by Greenhalgh *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Health Services and Delivery Research Editor-in-Chief

Professor Jo Rycroft-Malone Professor of Health Services and Implementation Research, Bangor University, UK

NIHR Journals Library Editor-in-Chief

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the EME Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Professor Aileen Clarke Professor of Public Health and Health Services Research, Warwick Medical School, University of Warwick, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Health Sciences Research, Health and Wellbeing Research Group, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Director, Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: nihredit@southampton.ac.uk

Abstract

Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care

Joanne Greenhalgh,^{1*} Sonia Dalkin,² Kate Gooding,¹
Elizabeth Gibbons,³ Judy Wright,¹ David Meads,¹ Nick Black,⁴
Jose Maria Valderas⁵ and Ray Pawson¹

¹School of Sociology and Social Policy, University of Leeds, Leeds, UK

²Department of Public Health, Northumbria University, Newcastle upon Tyne, UK

³Nuffield Department of Population Health, University of Oxford, Oxford, UK

⁴Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

⁵Institute of Health Research, University of Exeter, Exeter, UK

*Corresponding author j.greenhalgh@leeds.ac.uk

Background: The feedback of patient-reported outcome measures (PROMs) data is intended to support the care of individual patients and to act as a quality improvement (QI) strategy.

Objectives: To (1) identify the ideas and assumptions underlying how individual and aggregated PROMs data are intended to improve patient care, and (2) review the evidence to examine the circumstances in which and processes through which PROMs feedback improves patient care.

Design: Two separate but related realist syntheses: (1) feedback of aggregate PROMs and performance data to improve patient care, and (2) feedback of individual PROMs data to improve patient care.

Interventions: Aggregate – feedback and public reporting of PROMs, patient experience data and performance data to hospital providers and primary care organisations. Individual – feedback of PROMs in oncology, palliative care and the care of people with mental health problems in primary and secondary care settings.

Main outcome measures: Aggregate – providers' responses, attitudes and experiences of using PROMs and performance data to improve patient care. Individual – providers' and patients' experiences of using PROMs data to raise issues with clinicians, change clinicians' communication practices, change patient management and improve patient well-being.

Data sources: Searches of electronic databases and forwards and backwards citation tracking.

Review methods: Realist synthesis to identify, test and refine programme theories about when, how and why PROMs feedback leads to improvements in patient care.

Results: Providers were more likely to take steps to improve patient care in response to the feedback and public reporting of aggregate PROMs and performance data if they perceived that these data were credible, were aimed at improving patient care, and were timely and provided a clear indication of the source of the problem. However, implementing substantial and sustainable improvement to patient care required system-wide approaches. In the care of individual patients, PROMs function more as a tool to support patients in raising issues with clinicians than they do in substantially changing clinicians' communication

practices with patients. Patients valued both standardised and individualised PROMs as a tool to raise issues, but thought is required as to which patients may benefit and which may not. In settings such as palliative care and psychotherapy, clinicians viewed individualised PROMs as useful to build rapport and support the therapeutic process. PROMs feedback did not substantially shift clinicians' communication practices or focus discussion on psychosocial issues; this required a shift in clinicians' perceptions of their remit.

Strengths and limitations: There was a paucity of research examining the feedback of aggregate PROMs data to providers, and we drew on evidence from interventions with similar programme theories (other forms of performance data) to test our theories.

Conclusions: PROMs data act as 'tin openers' rather than 'dials'. Providers need more support and guidance on how to collect their own internal data, how to rule out alternative explanations for their outlier status and how to explore the possible causes of their outlier status. There is also tension between PROMs as a QI strategy versus their use in the care of individual patients; PROMs that clinicians find useful in assessing patients, such as individualised measures, are not useful as indicators of service quality.

Future work: Future research should (1) explore how differently performing providers have responded to aggregate PROMs feedback, and how organisations have collected PROMs data both for individual patient care and to improve service quality; and (2) explore whether or not and how incorporating PROMs into patients' electronic records allows multiple different clinicians to receive PROMs feedback, discuss it with patients and act on the data to improve patient care.

Study registration: This study is registered as PROSPERO CRD42013005938.

Funding: The National Institute for Health Research Health Services and Delivery Research programme.

Contents

List of tables	xiii
List of figures	xv
Glossary	xvii
List of abbreviations	xxv
Plain English summary	xxvii
Scientific summary	xxix
Chapter 1 Introduction and overview	1
Background and rationale for the review	1
Definitions and policy context	1
Existing evidence	2
Aims and objectives	5
Public and patient involvement	6
Rationale and overview of methodology	6
Why realist synthesis?	6
What is realist synthesis?	7
How is realist synthesis conducted?	8
<i>Step 1: searching for and identifying programme theories</i>	8
<i>Step 2: focusing the review and selecting programme theories</i>	8
<i>Step 3: searching for empirical evidence</i>	9
<i>Step 4: quality appraisal and data extraction</i>	9
<i>Step 5: synthesis</i>	10
Structure of the report	10
<i>Review 1: a realist synthesis of the feedback of aggregate patient-reported outcome measures and performance data to improve patient care</i>	10
<i>Review 2: a realist synthesis of the feedback of individual patient-reported outcome measures data to improve patient care</i>	10
Chapter 2 Review methodology: feedback and public reporting of aggregate patient-reported outcome measures data	11
Searching for and identifying programme theories	11
<i>Inclusion criteria</i>	12
<i>Exclusion criteria</i>	12
Focusing the review and selecting programme theories	13
<i>Agreeing the focus of the review</i>	13
<i>Identifying and searching for abstract theories</i>	15
<i>Searching for empirical evidence and selection of studies</i>	16
<i>Data extraction, quality assessment and synthesis</i>	18
Chapter summary	24
Chapter 3 Feedback of aggregate patient-reported outcome measures data: programme theory elicitation	25
Introduction	25

Policy and programme history	25
<i>History of performance measurement and reporting in the NHS</i>	25
<i>Outcomes as indicators of NHS performance</i>	27
<i>The rise of consumer choice and information</i>	27
<i>The UK national patient-reported outcome measures programme</i>	29
<i>Outcome and patient-reported outcome measures collection in other NHS services</i>	29
<i>Financial rewards and sanctions in England</i>	32
Summary	33
Programme theories	33
How the intervention is intended to improve patient care	34
<i>Supporting patient choice</i>	34
<i>Accountability to stakeholders</i>	35
<i>Provider benchmarking and peer competition</i>	36
<i>The role of the media</i>	37
<i>Assessing the appropriateness of surgery</i>	38
<i>Summary of how patient-reported outcome measures are intended to improve patient care</i>	39
Why the intervention may not work as intended	39
<i>Does patient choice drive improvements in patient care?</i>	39
<i>Unintended or adverse consequences of performance data</i>	40
<i>Data credibility and attribution: do performance data reflect the quality of care?</i>	42
<i>Actionability</i>	46
Substantive theories of patient-reported outcome measures feedback	52
<i>Audit and feedback</i>	52
<i>Benchmarking</i>	53
<i>Public disclosure</i>	55
An overall programme theory to guide the review	56
Chapter 4 Feedback of aggregate patient-reported outcome measures and performance data: reviewing mechanisms	59
Introduction	59
Theory 1: the feedback of patient-reported outcome measures or performance data leads to improved patient care	59
<i>Theory 1a: patient-reported outcome measures feedback will lead to improved patient outcomes</i>	60
<i>Theory 1b: the public reporting of performance data leads to improved patient outcomes</i>	61
<i>Theory 1 summary</i>	62
<i>Reviewing mechanisms: how and why does performance feedback lead to improvements in patient care?</i>	62
Theory 2: feedback influences provider behaviour according to whether it is delivered privately (confidentially) or publicly	64
<i>Theory 2a: providers are motivated to respond to private/confidential feedback on performance</i>	64
<i>Theory 2b: public reporting of performance places additional pressure on providers to respond</i>	66
<i>Theory 2 summary</i>	68
<i>Explaining how public disclosure works</i>	69
Theory 3: public reporting of poor performance threatens providers' market share and provides a significant provider response	69
<i>Theory 3a: providers experience a change in their market share following public reporting; theory 3b: providers take steps to improve the quality of care because they are worried about potential threats to their market share</i>	70
<i>Theories 3a and 3b summary</i>	72

<i>Theory 3c: providers perceive that report cards pose a threat to their market share</i>	72
<i>Theory 3c summary</i>	74
Theory 4: providers perceive that report cards damage their professional or their hospital's reputation	74
<i>Theory 4a: providers perceive that report cards affect market share by damaging a hospital's reputation</i>	75
<i>Theory 4a summary</i>	78
<i>Theory 4b: media reports of hospital performance damage a hospital's reputation and affect their market share</i>	78
<i>Theory 4b summary</i>	81
Theory 5: providers respond through comparing themselves with their peers	81
<i>Hildon et al.</i>	81
<i>Greer</i>	82
<i>Davies</i>	82
<i>Theory 5 summary</i>	83
Theory 6: consumers choose hospitals on the basis of information about the quality of services	83
<i>Theory 6a: patients are willing to move hospitals in response to poor performance;</i>	
<i>theory 6b: patients choose hospitals on the basis of reports about service quality;</i>	
<i>theory 6c: patients respond to poor hospital performance when it is reported in the media</i>	84
<i>Theory 6 summary</i>	92
Chapter summary	92
<i>Theory 1</i>	92
<i>Theory 2</i>	92
<i>Theory 3</i>	93
<i>Theory 4</i>	93
<i>Theory 5</i>	93
<i>Theory 6</i>	93
Conclusions	94
Chapter 5 Feedback of aggregate patient-reported outcome measures and performance data: reviewing contexts	95
Theory 7: financial incentives and sanctions influence providers' responses to the public reporting of performance data	96
<i>Theory 7a: financial incentives accelerate and amplify the impact of public reporting and feedback of performance data on improvements to patient care</i>	97
<i>Theory 7a summary</i>	101
<i>Theory 7b: providers do not make improvements to patient care when no financial incentives are attached to performance feedback</i>	101
<i>Theory 7b summary</i>	103
<i>Theory 7c: financial incentives attached to the feedback of performance data can lead to 'tunnel vision'</i>	103
<i>Theory 7 summary</i>	106
Theory 8: the perceived credibility of performance data influences providers' responses to the feedback of performance data	106
<i>Theory 8a: the perceived credibility of the performance data influences providers' responses to performance data</i>	107
<i>Theory 8a summary</i>	110
<i>Theory 8b: the source of performance data influences providers' perceptions of their credibility</i>	110
<i>Theory 8b summary</i>	113
<i>Theory 8c: the perceived underlying driver of public reporting systems influence providers' responses</i>	113

<i>Theory 8c summary</i>	119
<i>Theory 8d: clinicians have greater trust in clinician-led reporting systems</i>	119
<i>Theory 8 summary</i>	123
Theory 9: the degree to which performance data are 'actionable' influences providers' responses to the feedback of performance data	123
<i>Theory 9a: patient experience data are actionable and enable providers to take steps to improve patient care</i>	124
<i>Theory 9b: making patient experience data more immediate and integrated into clinical discussions improves provider responses; theory 9c: providing tailored support to interpret and act on patient experience data improves providers' responses</i>	131
Chapter summary	135
<i>Theory 7</i>	135
<i>Theory 8</i>	136
<i>Theory 9</i>	136
Chapter 6 Review methodology: feedback of individual patient-reported outcome measures in the care of individual patients	139
Searching for and identifying programme theories	139
<i>Inclusion criteria</i>	139
<i>Exclusion criteria</i>	139
Focusing the review and selecting programme theories	139
<i>How patients might use patient-reported outcome measures</i>	141
<i>How clinicians might use patient-reported outcome measures</i>	141
<i>A review to compare patient-reported outcome measures use in different long-term conditions</i>	141
<i>A review of specific blockages in the patient-reported outcome measures feedback implementation chain</i>	142
<i>A review of the 'system strains' to patient-reported outcome measures feedback</i>	142
Searching for empirical evidence and selection of studies	143
<i>Inclusion criteria</i>	143
<i>Exclusion criteria</i>	143
Data extraction, quality assessment and synthesis	144
Chapter summary	145
Chapter 7 Candidate programme theories of how the feedback of patient-reported outcome measures is intended to improve the care of individual patients	147
Patient-reported outcome measures in the care of individual patients: a brief programme history	147
Candidate programme theories underlying patient-reported outcome measures feedback in the care of individual patients	149
<i>Patient-reported outcome measures as screening tools</i>	151
<i>Patient-reported outcome measures as clinical monitoring tools</i>	151
<i>Patient-reported outcome measures as a tool to support personalised care planning and self-management</i>	152
Counter programme theories: how the feedback of patient-reported outcome measures may not lead to the intended outcomes of improving patient care	158
<i>Patient-reported outcome measures feedback is not practical or feasible</i>	159
<i>Psychometric properties of current patient-reported outcome measures do not support their use in clinical practice</i>	159
<i>Patient-reported outcome measures may constrain rather than support the clinician–patient relationship</i>	160
<i>The challenges of using patient-reported outcome measures for multiple purposes</i>	162
An overall programme theory	162

Chapter 8 Patient-reported outcome measures as a tool to support patients in raising or sharing concerns with clinicians	165
Introduction	165
Patient-reported outcome measures as a tool to support clinical decision-making independently of the patient consultation	165
Patient-reported outcome measures as a tool to promote patient self-management of long-term conditions	166
Patient-reported outcome measures as a tool to facilitate patient self-reflection and to help patients raise or share issues with clinicians	166
Patient-reported outcome measures as tool for raising clinicians' awareness of patient concerns	166
Summary and structure of our review	167
Candidate programme theories: patient-reported outcome measures as a tool to enable patients to raise or share issues with clinicians	167
Evidence review: theory testing and refinement	169
<i>Patients' views on the structure and format of patient-reported outcome measures</i>	170
Qualitative evaluations of standardised and individualised patient-reported outcome measures	175
Mallinson	175
Westerman et al.	176
Farquar et al.	177
<i>Summary of qualitative findings on patient-reported outcome measure format and structure</i>	178
Reviewing context configurations: how do they shape the intended impact of patient-reported outcome measures completion?	179
<i>Patients' and clinicians' experiences of using patient-reported outcome measures to detect and manage depression in primary care</i>	179
<i>Patients' and clinicians' experiences of using patient-reported outcome measures in secondary care mental health services</i>	185
<i>Clinicians' and patients' views of patient-reported outcome measures in palliative care</i>	189
Chapter summary	201
Theory 10	202
Theory 11	202
Theory 12	202
Theory 13	203
Theory 14	203
 Chapter 9 Patient-reported outcome measures as a tool for raising clinicians' awareness of patients' concerns	 205
Introduction	205
Candidate programme theories	206
Evidence review	206
Velikova et al.	207
Velikova et al.	208
Takeuchi et al.	209
Mills et al.	210
Detmar et al.	211
Berry et al.	212
Summary	213
Theories 15 and 16	213
Theory 17	214

Further theory testing: why does patient-reported outcome measures feedback not change clinicians' communication practices?	214
<i>Theory 18</i>	215
<i>Theory 19</i>	215
Chapter summary	224
<i>Theory 18</i>	224
<i>Theory 19</i>	225
Chapter 10 Discussion and conclusions	227
Summary of main findings	227
Chapter 3: <i>a taxonomy of the programme theories underlying the use of aggregate patient-reported outcome measures data</i>	227
Chapter 4: <i>the mechanisms through which the feedback of patient-reported outcome measures and performance data works</i>	228
Chapter 5: <i>contextual configurations that influence the mechanisms through which performance feedback works and the resulting outcomes</i>	229
Chapter 7: <i>programme theories underlying the use of individual patient-reported outcome measures data to support the care of patients</i>	230
Chapter 8: <i>patient-reported outcome measures as a tool to support patients in raising or sharing concerns with clinicians</i>	230
Chapter 9: <i>patient-reported outcome measures as a tool for raising clinicians' awareness of patients' concerns</i>	231
Strengths and weaknesses of our review	232
Discussion and implications for practice	233
<i>Feedback of aggregate-level patient-reported outcome measures data</i>	233
<i>The use of financial incentives</i>	234
<i>Credibility</i>	235
<i>Actionability</i>	235
<i>The feedback of patient-reported outcome measures data in the care of individual patients</i>	236
<i>Collection of patient-reported outcome measures for multiple purposes?</i>	237
Recommendations for future research	237
Acknowledgements	241
References	243
Appendix 1 Search strategies	267
Appendix 2 Example of a data extraction form	271
Appendix 3 Example of one 'theory theme'	273
Appendix 4 Table of theories tested and included studies	275
Appendix 5 Dissemination plan	279

List of tables

TABLE 1 Summary of initial searches and included studies for aggregate PROMs feedback evidence review	18
TABLE 2 Indicators and source of data provided for each hospital on the NHS Choices website	28
TABLE 3 The ISOQOL's minimum standards for the use of PROs in patient-centred and comparative effectiveness research	148
TABLE 4 Different types of PROMs	150
TABLE 5 Patient ratings for the SF-36 and EQ-5D	173
TABLE 6 A comparison of the characteristics of different public reporting of performance initiatives in the English NHS	234
TABLE 7 Future research priorities	239

List of figures

FIGURE 1 Summary of searching and selection process for aggregate programme theories	13
FIGURE 2 Summary of abstract theory searches	16
FIGURE 3 Flow diagram of the evidence search process for the feedback of aggregate PROMs data	19
FIGURE 4 Mind map of the process through which providers respond to performance data	21
FIGURE 5 Original and a revised logic model of provider responses to performance data following feedback of 'poor' performance	22
FIGURE 6 Sequence of 'middle range' theories about how and why providers respond to performance data	23
FIGURE 7 NHS Outcomes Framework	27
FIGURE 8 Flow chart for the collection of PROMs data in the English national PROMs programme	30
FIGURE 9 An example of a funnel plot	31
FIGURE 10 van Helden and Tilemma's benchmarking model	54
FIGURE 11 An overall programme theory to guide the review	57
FIGURE 12 Sources of information used by patients offered a choice of hospital	85
FIGURE 13 Most importance factors for patients offered a choice when choosing a hospital	86
FIGURE 14 Factors that influenced patients' choice of hospital	90
FIGURE 15 Contextual complexity and outcomes	95
FIGURE 16 Theory searches for individual PROMs feedback	140
FIGURE 17 Process of study review and selection for the individual-level synthesis	144
FIGURE 18 Greenhalgh <i>et al.</i> 's model of PROMs feedback in the care of individual patients	153
FIGURE 19 Model of the consultation in collaborative personalised care planning	154
FIGURE 20 How PROMs completion can support the management of people with long-term conditions	155

FIGURE 21 Etkind <i>et al.</i> 's logic model of PROMs feedback in palliative care	157
FIGURE 22 Street <i>et al.</i> 's model of the direct and indirect pathways from communication to outcomes	158
FIGURE 23 Patient-reported outcome measures feedback in the care of individual patients: a logic model	163

Glossary

Agency for Healthcare Research and Quality A subdivision of the US Department of Health and Human Services, focused on enhancing the quality, appropriateness and effectiveness of health-care services and access to care.

Annual Healthcheck A system of rating the quality of hospitals in England between 2005 and 2009, which was overseen by the Healthcare Commission.

Attention Deficit Hyperactivity Disorder A group of behavioural symptoms, including inattentiveness, hyperactivity and impulsiveness, which tend to be identified at an early age.

Best Practice Tariff A system of payment introduced in England to pay prices for hospital services that reflect the cost of best clinical practice rather than pay the national average cost of care.

British Medical Journal A weekly peer-reviewed medical journal published by the BMJ Group.

California Hospital Outcomes Project A system of annual public reports on risk-adjusted hospital outcomes for medical, surgical and obstetric patients in California, overseen by a government agency, the Office of State Wide Health Planning and Development.

Cardiac Care Network A support system to the Ministry of Health and Long-Term Care that is focused on improving quality, efficiency, access and equity in the delivery of the continuum of adult cardiac services in Ontario, Canada.

Care management practices Organised processes implemented by clinician groups to methodically improve the quality of care for patients.

Care Quality Commission The health and social care regulator in England.

Centers for Medicare and Medicaid Services A US federal agency that administers government-organised health insurance schemes, including Medicare, Medicaid and the State Children's Health Insurance Program.

Chief executive officer The most senior manager in an organisation.

Child and adolescent mental health services Mental health services for children and adolescents in England.

Children and Young People's Improving Access to Psychological Therapy In 2011, the English government launched a strategy, 'No Health Without Mental Health', to improve the care and services for adults and children with mental health problems. Part of this strategy focused on providing additional funding to increase access to psychological therapies that were deemed effective by the National Institute for Health and Care Excellence, using a stepped-care approach.

Choice-based conjoint analysis A form of statistical analysis used to analyse discrete choice experiments in which patients are asked to choose between a number of scenarios/options that systematically vary on a number of different dimensions (e.g. quality, convenience or waiting time).

Chronic obstructive pulmonary disease The collective name for a number of lung diseases, including bronchitis, emphysema and chronic obstructive airways disease, all characterised by breathing problems.

Client Generated Index An individualised patient-reported outcome measure.

Clinical Commissioning Group An organisation set up by the Health and Social Care Act 2012 to organise and commission the delivery of NHS services in England.

Clinical Resource and Audit Group Set up by the Scottish Health Executive in 1989 as a subgroup of the Clinical Effectiveness Strategy Group. Its aim was to support clinicians to examine and improve the quality of their care.

Commission for Health Improvement The hospital regulator in England between 2000 and 2004.

Commissioning for Quality and Innovation National and local quality improvement goals and associated financial rewards for providers in England.

Consumer Assessment of Healthcare Providers and Systems A programme funded by the Agency for Healthcare Research and Quality to administer, collect and report on surveys asking patients to evaluate and report on their experiences of health services.

Context–mechanism–outcome configuration A heuristic used in realist evaluation and synthesis to express theories about how context shapes the mechanism through which a programme works and thus its subsequent outcome(s).

Contextual Feedback Intervention Theory A theory that seeks to explain how feedback may work to prompt clinicians to change their behaviour.

Coronary artery bypass surgery Heart surgery in which a blood vessel is redirected to bypass a blocked blood vessel to enable blood to flow around the heart.

Coronary care unit A specialist unit in a hospital caring for patients with cardiac conditions such as heart attack and unstable angina.

Disease Repercussion Profile An individualised patient reported outcome measure developed for use in rheumatology.

District General Hospital A hospital providing services to a local community that is not linked to a medical school.

Electronic Medical Record An electronic record storing information about a patient's medical history, treatment and episodes of care.

Electronic patient-reported outcome measure A patient-reported outcome measure that is completed electronically.

Electronic Self-Report Assessment of Cancer A web-based patient-reported outcome measure designed to assess the impact of cancer on quality of life.

European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core –30 A questionnaire designed to assess the quality of life of cancer patients. It can be supplemented with cancer site-specific modules.

EuroQol 5D A utility-based patient-reported outcome measure.

Feedback Intervention Theory A theory that sets out how individuals are expected to respond to feedback.

Forced expiratory volume A measure of how much air a person can exhale in one breath, which is often used as a measure of the severity of lung diseases, including asthma.

Friends and Family Test Launched in 2013 by NHS England to provide managers and clinicians with rapid feedback on services from patients. Patients are asked to indicate, on a five-point scale from 'extremely likely' to 'extremely unlikely', 'how likely are you to recommend our service to friends and family if they needed similar care or treatment?'

Functional Assessment of Cancer Therapy – General A questionnaire designed to measure the quality of life of people with cancer, consisting of four subscales measuring physical, social/family, emotional and functional well-being. Can be supplemented with cancer site-specific subscales.

Functional Assessment of Cancer Therapy – Lung A questionnaire designed to measure the quality of life of people with lung cancer, consisting of four subscales from the Functional Assessment of Cancer Therapy – General plus an additional subscale focusing on the concerns of patients with lung cancer.

General Health Questionnaire A questionnaire designed to screen for minor psychiatric disorders in the general population.

General practitioner A doctor who works in primary care and provides generalist care, treatment and preventative services for a range of chronic and acute illnesses.

Health and Social Care Information Centre A national provider of information, data and information technology systems for health and social care in England.

Health Care Financing Administration Currently known as Centers for Medicare and Medicaid Services, a federal agency in the USA set up in 1977 to oversee the administration of the Medicare programme (medical insurance) and quality assurance activities, including the publication of mortality report cards in 1987.

Health-care provider A person or organisation providing health care.

Health Maintenance Organisation A medical insurance group in the USA that provides medical insurance on a fixed, pre-paid basis.

Health-related quality of life A contested concept with a number of definitions; most definitions relate to the idea that it is a multidimensional concept reflecting a patient's perceptions of their functioning across a range of domains (e.g. physical, social and emotional).

Heart failure A condition caused by the failure of the heart to pump blood around the body with sufficient pressure because the heart muscles are too stiff or do not work as they should.

Hip Osteoarthritis Outcome Score A patient-reported outcome measure developed for people with osteoarthritis.

Hospice Quality of Life Index A patient-reported outcome measure designed for use in palliative care settings.

Hospital Anxiety and Depression Scale A questionnaire, designed to screen for anxiety and depression, for use in community and hospital populations.

Hospital Episode Statistics Data collected on all admissions, outpatient appointments and accident and emergency attendances at NHS hospitals in England.

Hospital Quality Alliance One of the first national initiatives in the USA to routinely publicly report on the care provided in all hospitals from 2004 onwards.

Hospital Quality Incentive Demonstration A US pay-for-performance scheme set up in 2003, in which hospitals were awarded bonus payments for high quality in several clinical areas, alongside a system of publicly reporting performance.

Hospital Quality Initiative A voluntary public report system launched by the Centers for Medicare and Medicaid Services in the USA.

Institute for Clinical Evaluative Sciences A not-for-profit research institute in Canada that carries out research and produces outputs based on Ontario's health-related data.

Intensive care unit A specialist hospital unit providing care for critically ill patients.

International Society for Quality of Life Research A non-profit international society established in 1993 to advance the scientific study of health-related quality of life and other patient-centred outcomes.

Intraclass correlation A measure of the reliability of two assessments or ratings.

Item response theory A paradigm for the design, analysis and scoring of patient-reported outcome measures.

Joint Commission on Accreditation of Healthcare Organizations An independent, not-for-profit organisation that is responsible for accrediting and certifying health-care organisations in the USA.

Massachusetts Health Quality Partners A collaborative of physicians, hospitals, health plans, purchasers, patient and public advocates, government agencies and academics, whose aim is to improve the quality of patient care through publicly reporting performance.

McGill Quality of Life Questionnaire A measure of quality of life designed for use in palliative care settings.

Measure Yourself Medical Outcome Questionnaire An individualised patient-reported outcome measure originally designed and tested for use in general practice.

Medical Care Questionnaire A questionnaire designed to measure patients' perceptions of the continuity and co-ordination of medical care.

Meticillin-resistant *Staphylococcus aureus* A type of bacteria that patients in hospital are at risk of becoming infected with and that can be resistant to antibiotics and, thus, can be difficult to treat.

Minimal clinically important difference The smallest difference in score in a subscale of a patient-reported outcome measure that patients consider to be beneficial and would warrant a change in the patient's management.

Montgomery-Åsberg Depression Rating Scale A depression rating scale designed to be sensitive to changes following treatment for depression.

Myocardial infarction The medical name for a heart attack that occurs when blood flow to the heart is blocked.

National Health Service The publicly funded health service in the UK.

National Institute for Health and Care Excellence An executive non-departmental public body of the Department of Health in England with responsibility for reviewing evidence and publishing guidance on effective care and treatment.

National Joint Registry Set up by the Department of Health and Welsh Government in 2002 to collect information on all hip, knee, ankle, elbow and shoulder replacement operations, and to monitor the performance of joint replacement implants and the effectiveness of different types of surgery.

National Study of Small and Medium Sized Physician Practices A national survey conducted in the USA to provide information about physician practices with 1–19 physicians, focusing on practices' use of information technology and organised care management processes.

New York State Cardiac Reporting System A reporting system in New York, initiated in 1989, as a collaborative venture between the New York State Department of Health Authority and 21-member appointed Cardiac Advisory Committee to produce cardiac surgery mortality report cards derived from chart data.

Nottingham Health Profile A patient-reported outcome measure designed to measure how people feel about various states of ill-health across a range of chronic conditions.

Office of State Wide Health Planning and Development Collects and disseminates information about California's health-care infrastructure. Oversaw the California Hospital Outcomes Project, which publicly reported mortality rates following cardiac surgery in California.

Oxford Hip Score A patient-reported outcome measure developed to assess function and pain in patients undergoing total hip replacement surgery.

Palliative Care Quality of Life Index A patient-reported outcome measure designed to assess the quality of life of patients with terminal cancer receiving palliative care.

Palliative Outcome Scale A patient-reported outcome measure designed for use in palliative care settings.

Patient Health Questionnaire A questionnaire designed to screen and monitor the severity of depression that can be used in clinical practice.

Patient-led Assessment of the Care Environment A system introduced into the English NHS in 2013 for assessing the quality of the patient environment in hospitals conducted by trained patients.

Patient Reported Outcomes: Feedback Interpretation and Learning A trial that explored whether or not providing benchmarked patient-reported outcome measures feedback to surgeons would improve patient outcomes.

Patient-reported experience measure A questionnaires that measures patients' experiences of care.

Patient-reported outcome An outcome directly reported by patients. In the USA, this term is often used in place of patient-reported outcome measure.

Patient-reported outcome measure A questionnaire that measures patients' perceptions of the impact of a condition and its treatment on their health.

Patient-Reported Outcome Measurement Information System A system of measures of patient-reported health status for physical, mental and social well-being, developed using item response theory.

Payment by Results A payment system in England under which commissioners pay health-care providers for each patient seen or treated, taking into account the complexity of each patient's health-care needs.

Percutaneous coronary intervention A non-surgical procedure used to treat people experiencing a heart attack or unstable angina, whereby a thin catheter is used to insert a stent in order to widen the artery and enable blood to flow around the heart more effectively.

Performance Assessment Framework A performance monitoring system that was introduced in England in 1999. Each hospital in England was measured against 60 indicators, which covered health improvement, fair access, effective delivery of appropriate health care, efficiency, patient/carer experience and health outcomes, to produce a 'balanced scorecard', that was made publicly available.

Performance indicator A measure designed to quantify the performance of an organisation.

Physician Organisation of Care Management Index A study-specific summary measure (ranging from 0–24) of whether the practice uses reminders for preventative care and doctor feedback, has a disease register, has clinical practice guidelines and employs non-doctor staff educators in the care of people with four chronic conditions (heart failure, depression, asthma and diabetes).

Preferences and Perceptions of Communication A questionnaire to measure patients' preferences and perceptions of communication during consultations.

Primary care trust An administrative body responsible for commissioning primary, secondary and community health services in the English NHS between 2001 and 2013.

Public and patient involvement The involvement of patients and the public in the design, management and conduct of research or the design of health services.

Public reporting Making performance data available to a range of stakeholders, including the general public.

Quality and Outcomes Framework A system of financially rewarding general practitioners in the English NHS for fulfilling a set of performance indicators designed to represent high-quality care.

Quality improvement A formal approach to the analysis of performance and systematic efforts to improve it.

Quality Improvement Organization A programme in the USA overseen by the Centers for Medicare and Medicaid Services, whereby a group of health quality experts, clinicians and consumers work together to improve the care delivered to people with Medicare.

Quality of life A highly contested concept with numerous definitions. The World Health Organization defines it as 'an Individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns' (source: World Health Organization; www.who.int/healthinfo/survey/whoqol-qualityoflife/en/ (accessed 10 October 2016)).

Randomised controlled trial A methodology for evaluating the effectiveness of an intervention through randomly allocating participants to control and intervention groups and measuring the change in outcome from pre to post intervention.

Realist And Meta-narrative Evidence Syntheses: Evolving Standards A project that developed reporting standards for two review methodologies: realist synthesis and meta-narrative review.

Schedule for the Evaluation of Individual Quality of Life An individualised patient-reported outcome measure administered using a semistructured interview.

Schedule for the Evaluation of Individual Quality of Life Direct Weight An individualised patient-reported outcome measure administered using a semistructured interview and where individuals determine the relative importance of each domain using a direct weighting procedure involving manipulating the size of five laminated disks.

Short Form Health Survey A patient-reported outcome measure designed for use in patients with chronic illness and healthy populations.

Sickness Impact Profile A patient-reported outcome measure designed to provide a generic measure of health status.

Society for Cardiothoracic Surgery A representative body for cardiothoracic surgery in Britain and Ireland.

Specialist registrar A doctor in the Republic of Ireland or the UK who is receiving specialist training. In the UK these posts have now been replaced by specialty registrars.

Supportive Care Needs Survey-Short Form A patient-reported outcome measure designed to assess the care needs of people with cancer.

Symptoms and Concerns Checklist A questionnaire designed to structure the ways in which palliative care nurses conduct their initial assessment of patients.

Trial Outcome Index A summary measure consisting of the sum of the physical well-being and functional well-being subscales from the Functional Assessment of Cancer Therapy – General and the ‘additional concerns’ subscales of the Functional Assessment of Cancer Therapy – Lung.

Visual analogue scale A form of response option in a questionnaire whereby respondents indicate (for example) the severity of pain by placing a mark on a continuous line between two end points representing extremes.

Wisconsin Collaborative for Healthcare Quality A collaborative founded in 2003 by chief executives of several large multispecialty practices and their partner hospitals in Wisconsin, USA, to promote quality improvement.

List of abbreviations

BMJ	<i>British Medical Journal</i>	HRQoL	health-related quality of life
BPT	Best Practice Tariff	HSCIC	Health and Social Care Information Centre
CAHPS	Consumer Assessment of Healthcare Providers and Systems	IAPT	Improving Access to Psychological Therapies
CAMHS	child and adolescent mental health services	ISOQOL	International Society for Quality of Life Research
CEO	chief executive officer	IT	information technology
CGI	Client Generated Index	JCAHO	Joint Commission on Accreditation of Healthcare Organizations
CHI	Commission for Health Improvement	MCQ	Medical Care Questionnaire
CHOP	California Hospital Outcomes Project	MHQP	Massachusetts Health Quality Partners
CMP	care management practice	MQOL	McGill Quality of Life Questionnaire
CMS	Centers for Medicare and Medicaid Services	NICE	National Institute for Health and Care Excellence
CQC	Care Quality Commission	NJR	National Joint Registry
CQUIN	Commissioning for Quality and Innovation	NYSCRS	New York State Cardiac Reporting System
CRAG	Clinical Resource and Audit Group	OHS	Oxford Hip Score
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30	PCOMI	Physician Organisation of Care Management Index
ePRO	electronic patient-reported outcome	PCT	primary care trust
EQ-5D	EuroQoL-5 Dimensions	PHQ-9	Patient Health Questionnaire
FACT-G	Functional Assessment of Cancer Therapy – General	POS	Palliative Outcome Scale
FACT-L	Functional Assessment of Cancer Therapy – Lung	PPI	public and patient involvement
FFT	Friends and Family Test	PREM	patient-reported experience measure
GP	general practitioner	PRO	patient-reported outcome
HCFA	Health Care Financing Administration	PROFILE	Patient Reported Outcomes: Feedback Interpretation and Learning
HQA	Hospital Quality Alliance	PROM	patient-reported outcome measure
HQID	Hospital Quality Incentive Demonstration	PROMIS	Patient Reported Outcome Measurement Information System
HQLI	Hospice Quality of Life Index	QI	quality improvement
		QOF	Quality and Outcomes Framework

LIST OF ABBREVIATIONS

RAMESES	Realist And Meta-narrative Evidence Syntheses: Evolving Standards	SEIQoL-DW	Schedule for the Evaluation of Individual Quality of Life Direct Weight
RCT	randomised controlled trial	SF-36	Short Form Questionnaire-36 items
SCC	Symptoms and Concerns Checklist	SIP	Sickness Impact Profile
SCNS-SF-34	Supportive Care Needs Survey-Short Form	SQLI	symptoms and quality of life issues
SEIQoL	Schedule for the Evaluation of Individual Quality of Life	WCHQ	Wisconsin Collaborative for Healthcare Quality

Plain English summary

Asking patients to complete questionnaires about the impact of their condition and its treatment on their health (known as patient-reported outcome measures, or PROMs) and giving the results to hospitals and clinicians is thought to help improve patient care. We carried out two literature reviews to explore this.

We searched electronic databases and checked the reference lists of existing literature reviews to find relevant papers. We consulted with a patient and a stakeholder group to agree the focus of the two reviews. We extracted relevant information from these papers and compared the findings across them to see whether or not PROMs feedback worked as expected, and, if not, why not.

Patients do not use information on hospital quality to choose a hospital but instead rely on their own experience and that of family and friends. Hospitals use this information to improve patient care if they trust the accuracy of the results, if they feel that the questionnaires are being collected to improve patient care and if the results tell them what the problems are. PROMs are useful to enable patients with long-term conditions to raise or share their concerns with doctors, but do not always change what doctors ask patients about during consultations. Doctors have some concerns that PROMs may raise issues that they do not feel trained to address or do not know how to treat. Future work should examine whether or not it is possible to collect PROMs data to support the care of individual patients and to improve the quality of services at the same time.

Scientific summary

Background

Patient-reported outcome measures (PROMs) are questionnaires that measure patients' perceptions of the impact of a condition and its treatment on their health. The national PROMs programme was introduced in the NHS with the intention of supporting patient choice, provider accountability and provider benchmarking, and, thus, improving patient care. Alongside the use of PROMs data at an aggregate level, the routine collection and use of PROMs data at the individual patient level has become more widespread, although in a less co-ordinated way. At the individual level, the intention of PROMs feedback is to improve the detection of patient problems, to support clinical decision-making about treatment through ongoing monitoring and to empower patients to become more involved in their care. However, reviewing the literature in this area is challenging owing to the complexity and heterogeneity of the intervention, as well as the variations in context into which this intervention is implemented. Our project aimed to address these challenges and to carry out a realist synthesis of the processes through which and circumstances in which PROMs feedback improves patient care.

Objectives

1. Identify and classify the various ambitions for the use of PROMs data at the aggregate and individual level to:
 - i. produce a comprehensive taxonomy of the 'programme theories' underlying these different functions, and capture their subtle differences and the tensions between them
 - ii. produce a logic model of the organisational logistics, social processes and decision-making sequences that underlie the collation, interpretation and utilisation of PROMs data.
2. Test and refine these programme theories about how PROMs feedback is supposed to work against existing evidence of how it works in practice to:
 - i. identify the implementation processes that support or constrain the successful collation, interpretation and utilisation of PROMs data
 - ii. identify the mechanisms and circumstances through which the unintended consequences of PROMs data arise and those in which they can be avoided.

Methods

We conducted two separate but related realist reviews that explored (1) the processes through which, and circumstances in which, the feedback of aggregate PROMs and performance data leads to providers taking steps to improve patient care; and (2) the processes through which, and circumstances in which, PROMs act as a tool (a) to support patients in raising or sharing their concerns with clinicians, and (b) for raising clinicians' awareness and discussion of patients' concerns.

For both reviews, we utilised two search strategies across several electronic databases to identify policy documents, opinion pieces, letters, commentaries, editorials and reviews that discussed how PROMs and (for the aggregate review) other performance data were intended to improve patient care. We developed a comprehensive taxonomy of the underlying ideas and assumptions, or *programme theories*, about how PROMs feedback is intended to work. We verified and extended these ideas with our patient group and a

stakeholder group of NHS clinicians and policy-makers. The project group subsequently agreed the focus of the two reviews.

For the review of aggregate PROMs and performance data, we worked on information specialist-designed search strategies to identify papers that explored providers' experiences of using, and responses to, PROMs and other performance feedback (mortality report cards, other 'performance data' and patient experience measures). We searched a number of electronic databases, including EMBASE Classic+EMBASE (via Ovid), Health Management Information Consortium (via Ovid), (Ovid) MEDLINE® and (Ovid) MEDLINE® In-Process & Other Non-Indexed Citations. We carried out backwards and forwards citation tracking of key systematic reviews. We developed a set of inclusion and exclusion criteria to identify papers relevant to testing our programme theories; following several iterative screening procedures, 58 papers were included in the final synthesis.

For the individual review, we used backwards citation tracking of six key papers, which identified 372 papers, and forwards citation tracking of five key papers, which identified 605 papers. We developed a set of inclusion and exclusion criteria, and 36 papers were included in the final synthesis.

For each review, we developed a logic model of the intended processes through which PROMs feedback was intended to improve patient care. We matched our programme theories to each stage in these logic models, and they served as a framework for the review. For each review, we conducted a pilot synthesis on a purposively selected sample of papers to refine our theories and then conducted the synthesis across all papers. The aggregate review tested and refined nine theories that focused on understanding the mechanisms through which providers were expected to respond to PROMs and other performance data, and the circumstances in which these data led to providers taking steps to improve patient care. For the individual review, we tested and refined eight theories that examined the processes through which, and circumstances in which, PROMs feedback (1) enabled patients to raise or share concerns with clinicians, and (2) raised clinicians' awareness of patients' concerns and led to discussion during the consultation. For both reviews, we assessed the quality of, and extracted data for, aspects of the study relevant to testing our theories. To aid the process of synthesis, we organised the studies by theory and completed data extraction tables to allow both cross-study and within-study analysis. The synthesis team (JG, SD, KG and EG) held regular meetings to discuss the findings, and discussed emerging findings with the wider project team. Ongoing findings were also discussed with our patient group.

Results

Feedback and public reporting of aggregate patient-reported outcome measures and performance data

Public reporting places additional pressure on providers, particularly poor performers, to respond. Providers perceive that the public reporting of poor performance damages their reputation, so they take action to improve patient care in response to this. Patients do not use publicly reported information about service quality to inform their choice of hospital, but instead rely on their personal experience, the opinions of friends and family and advice from their general practitioner. Providers perceive mandatory public reporting programmes initiated by regulators or national or state governments as being driven by political motives, while those initiated by employers or insurance companies were perceived as being driven by a desire to cut costs or increase a provider's market share. Under these circumstances, providers criticised data about service quality as lacking credibility because they were based on data designed for a different purpose, had inadequate methods of case-mix adjustment and did not reflect what was clinically important. When performance data were fed back privately, with no public reporting, providers either ignored these data or attempted to improve data collection practices. When these data were publicly reported, they led providers to focus on improving those areas of care subjected to measurement at the expense of other areas of care: so-called 'tunnel vision' or 'effort substitution'.

Providers responded to performance data when their own internally collected data also suggested that there was a problem with the quality of care. Furthermore, externally mandated publicly reported data, particularly data focusing on outcomes, did not enable providers to identify the causes of any areas of poor care. Rather, additional investigations were needed, which required additional resources and the 'know-how' and capacity to carry out these investigations. Clinically initiated public reporting systems were perceived as driven by a desire to improve the quality of patient care, which secured clinical involvement. Clinical involvement in these programmes ensured that the indicators represented areas of care that clinicians perceived as important, and clinical ownership of the sources of data, indicator specification and methods of case-mix adjustment meant that it was difficult for clinicians to dismiss or ignore these data. Under these circumstances, providers took steps to improve patient care in order to be as good as or better than their peers, and did so through sharing and learning from best practices. However, action depended on the providers' experience of quality improvement (QI): providers with more experience were more likely to make sustainable improvements.

Providers valued data that were timely and specific, and that provided a clear indication of which care processes needed to be improved. However, the feedback of these data did not always lead to providers taking steps to improve patient care. When change did occur, efforts were more likely to be directed at less complex, discrete organisational aspects of care. Changes that required clinicians to modify the interpersonal aspects of their care, or that were more complex, were perceived as more difficult to implement. Furthermore, changes to one aspect of patient care could have unintended effects on other aspects of care. Significant and sustained improvements in patient care in response to the feedback of performance data can be achieved only through system- and organisation-wide strategies.

Patient-reported outcome measures in the care of individual patients

Whether PROMs support or constrain patients in sharing or raising issues with clinicians depends on the structure of the PROM. Standardised PROMs were useful for those patients who preferred not to talk about personal or sensitive issues, helping them to share information. However, clinicians in primary and secondary mental health settings and palliative care perceived that standardised PROMs constrained the patient-clinician relationship because they trivialised patients' emotions or did not capture the complex and dynamic nature of patients' problems. When there were no incentives attached to the use of standardised PROMs, clinicians avoided using them because they did not support the care of patients. In some situations, clinicians also adapted or changed the PROM to make it more useable, which may have compromised its validity as an instrument to support the care of individual patients. When clinicians were financially incentivised to use standardised PROMs, they used a wide range of tactics to adapt the PROMs to fit their interactions with patients in order to avoid being penalised for avoiding PROMs use. Some of these strategies may have compromised the validity of the PROM as tool to support the care of patients and as an indicator of the quality of care.

Clinicians and patients perceived that, when used in first assessments, individualised PROMs supported relationship-building because they enabled the patient to 'tell their story'. However, individualised PROMs were less useful as an outcome measure to judge change over time, owing to differences in the way that cues were defined between patients, between patients and interviewers, and over time. PROMs feedback could increase discussion of symptoms during the consultation and, in one study, led directly to improvements in patient well-being. The mechanism underlying this process is that patients felt more comfortable raising both physical and psychosocial or non-medical issues with their doctors and were more likely to initiate discussion about these during the consultation.

Patient-reported outcome measures do not substantially change doctors' communication practices during the consultation. Consultations still focused on symptoms rather than psychosocial issues, and clinicians were not more likely to initiate discussions about the latter. This was because doctors see their remit as dealing with issues specifically related to the patient's condition and its treatment, and consider that it is nurses' role to address wider psychosocial issues, a perception shared by nurses themselves. The limited effect on discussion also occurred because doctors closed down discussions about issues they felt unable

to treat. In palliative care settings, although nurses recognised that PROMs could raise issues that fell outside their remit and that they could not address, they recognised that ‘just discussing’ these issues could have therapeutic value for patients.

Strengths and limitations

There was a paucity of research examining the feedback of aggregate PROMs data to providers, and we drew on evidence from interventions with similar programme theories (other forms of performance data) to test our theories.

Conclusions

Our findings suggest that actively involving professional organisations in the process of agreeing indicator specification and case-mix adjustment is important to secure clinical engagement with performance feedback. Explaining how the process of case-mix adjustment is achieved in a way that clinicians can understand could improve their credibility among clinicians. There is little variation between providers in their performance in the national PROMs programme, suggesting that there may be other services and settings in which there is greater variation in practice that would benefit from the routine collection of PROMs data. PROMs data act as ‘tin openers’ rather than as ‘dials’. They classify providers as ‘statistical’ outliers, but this does not necessarily mean that providers are providing poor care, nor does it inform providers about the possible causes of any poor care. Providers are expected to conduct further investigations to identify alternative explanations for their outlier status and, if none are found, to explore the possible causes of their outlier status. Providers need more support and guidance on how to collect their own internal data, how to rule out alternative explanations for their outlier status and how to explore the possible causes of their outlier status. Further support and guidance are also needed to enable providers to integrate and interpret PROMs data in the context of other pieces of performance data, in order to access a bigger picture overview of their performance.

Our review highlighted the importance of considering how PROMs feedback in the care of individual patients affects not only the information-exchange and decision-making functions of the consultation, but also the relationship-building function. PROMs function more as a tool to support patients in raising issues with clinicians than they do in substantially changing clinicians’ communication practices with patients. In settings such as palliative care and psychotherapy, clinicians viewed individualised PROMs as useful in building rapport and supporting the therapeutic process. However, individualised PROMs are more time-consuming to complete, and it is not clear if such measures would be feasible to collect in more time-pressured settings such as primary care or outpatient appointments for patients with long-term conditions. Future research on the use of PROMs in the care of individual patients should focus on ensuring that the different relevant clinicians gain access to this information, so that issues can be addressed by the clinician with the appropriate remit, through integrating PROMs collection into patients’ electronic records.

Finally, using PROMs data to support both service-level QI and the care of individual patients is still a challenge. There is a need for mixed-methods studies to explore how differently performing providers (e.g. positive, negative and those at average levels) have used PROMs to identify the areas of care that could be improved, as well as how they have used PROMs data to guide improvements. There is also a need for future research to examine how providers have collected PROMs data to support both the care of individual patients and service QI.

Study registration

This study is registered as PROSPERO CRD42013005938.

Funding

Funding for this study was provided by the Health Services and Delivery Research of the National Institute for Health Research.

Chapter 1 Introduction and overview

Background and rationale for the review

In this chapter we provide an overview of the background of, and rationale for, the review. We then summarise the aims and objectives of the review, and provide an overview of our methodological approach: realist synthesis. We then outline the structure of the report.

Definitions and policy context

Patient-reported outcome measures (PROMs) are questionnaires that measure patients' perceptions of the impact of a condition and its treatment on their health.¹ Many of these measures were originally designed for use in research to ensure that the patient's perspective was integrated into assessments of the effectiveness and cost-effectiveness of care and treatment.² Over the last 5 years, the routine collection of PROMs data has played an increasing role in health policy in England, with the introduction of the national PROMs programme in the NHS. The original challenge that led the Department of Health to pilot the routine collection of PROMs data in 2007 was a demand management issue: to assess whether or not surgery for certain conditions was overutilised.³ However, more recent findings showing the gradient in use of these interventions by social class and ethnicity have led to calls for PROMs to be used to improve the equity of care.⁴ In 2008, the Darzi report⁵ called for the routine collection of PROMs data to benchmark provider performance, assess the appropriateness of referrals, support the payment of providers by results and support patient choice of provider. The public reporting of PROMs data to support the patient choice agenda was given further impetus in the 2010 government White Paper,⁶ which set out that 'Success will be measured . . . against results that really matter to patients' and that 'Patients . . . will have more choice and control, helped by easy access to the information they need about the best GPs [general practitioners] and hospitals' (© Crown Copyright; contains public sector information licensed under the Open Government Licence v3.0). Most recently, in the light of the Francis report,⁷ it is planned to introduce single aggregated ratings, and to develop ratings of hospital performance at department level to support public accountability and patient choice.⁸

Alongside the use of PROMs data at an aggregate level, the routine collection and use of PROMs data at the individual patient level has also become more widespread, although in a less co-ordinated way, with individual clinicians using PROMs on an ad hoc basis, often with little guidance.⁹⁻¹¹ At the individual level, the intention of PROMs feedback is to improve the detection of patient problems, to support clinical decision-making about treatment through ongoing monitoring and to empower patients to become more involved in their care.^{12,13}

Despite the fact that the ambitions for the usage of PROMs data have multiplied, PROMs research has focused on form rather than function. There is a substantial body of evidence on the psychometric properties of PROMs, but less attention has been given to clarifying the subsequent decisions and actions that the measures are intended to support.¹² For example, careful deliberation went into selecting the instruments for the UK PROMs programme and piloting the feasibility of their collection,¹⁴ but the precise mechanisms through which PROMs data will improve the quality of patient care for each of their intended functions have been less well articulated.¹² Furthermore, there are inherent tensions between the different uses of PROMs data that may influence how these data are collated and interpreted and, thus, the success of PROMs initiatives.^{15,16} For example, individualised measures, where the patient specifies the domains to be measured, may be more relevant to patients and, thus, better support patient involvement in their care than standardised measures.¹⁷ However, such personalised measures may lose their meaning when used at the group level, and thus may not be adequate reflections of the quality of patient care. Accordingly, there

is a significant need for research that clarifies the different functions of PROMs feedback and delineates more clearly the processes through which they are expected to achieve their intended outcomes.

Despite their relatively recent introduction to the NHS, the underlying reasoning about how PROMs data will be mobilised is familiar, and has a long and somewhat chequered history. For example, the use of aggregated PROMs data to benchmark provider performance and the public reporting of these data to inform consumer choice shares many of the assumptions and some of the drawbacks of other 'feedback' or 'public disclosure' interventions (e.g. hospital star ratings in the UK¹⁸ and surgical mortality report cards in the USA¹⁹). These interventions may improve patient care through a 'change' pathway [whereby providers initiate quality improvement (QI) activities to improve the quality of patient care] or a 'selection' pathway (whereby patients choose a high-quality hospital).²⁰ Evidence across a range of different forms of this intervention suggests that the public reporting of performance data results in improvements in performance in situations in which the named party is motivated to maintain their market share; the reporting occurs alongside other market sanctions (e.g. financial incentives); the public reporting carries intensive but controllable media interest; the disclosed data are unambiguous in classifying poor and high performers; and the reporting authority is trusted by those who receive the data.²¹ The evaluation and implementation of the public reporting and feedback of PROMs data will benefit from a careful review of the extent to which equivalent conditions apply to the impact of public dissemination of these data on the quality of patient care. At the individual level, PROMs feedback for detection and monitoring of patients' problems can be seen as an attempt to modify clinical judgement with encoded, standardised knowledge as part of the move towards scientific-bureaucratic medicine.²² The intention of increasing patient involvement in the consultation bears the hallmarks of other collaborative care interventions,²³ and much can be learned by reviewing common underlying mechanisms. As PROMs feedback is rolled out to other services and settings, it is vital that such cumulative evidence on parallel interventions informs future implementation.

Existing evidence

There are currently no systematic reviews examining the feedback of aggregate PROMs data to improve patient care. Boyce and Browne's²⁴ systematic review examined PROMs feedback in the care of individual patients and at the aggregate level. They found only one study²⁵ examining the use of aggregate PROMs data, in which physicians were randomised either to receive peer comparison feedback on the functioning of older patients in their care or to be told that the functioning of their older patients would be monitored. This study found no statistically significant differences in patient functional status between patients in the intervention and control groups. There are four reviews of the feedback of performance data.^{19,26–28} In general, these reviews found a small decline in mortality following public reporting after controlling for trends in a reduction of mortality; however, individual studies varied in their findings. For example, studies examining the impact of cardiac public reporting programmes on mortality rates found a variable picture: eight studies found a decrease in mortality rates over time,^{29–36} while another four studies^{37–40} found no changes in mortality rates over time. Similarly, although most studies examining the impact of public reporting on process indicators found an improvement in hospital quality, this varied from a 'slight' improvement to a 'significant' improvement. However, they also found little evidence that the public reporting of performance data stimulated changes in hospitals' market share, suggesting that patients may not change hospitals in response to the public reporting of quality data. We consider these reviews in more detail in *Chapter 4* of this report.

There are 16 reviews of the quantitative/randomised controlled trial (RCT) literature on the feedback of individual-level PROMs data^{24,41–56} and one review currently in progress.⁵⁷ There is also one review of qualitative studies⁵⁸ and four mixed-method reviews.^{59–62} Thus, there are a total of 21 existing reviews examining the feedback of individual PROMs data in patient care. Of these reviews, one focused on screening for mental health problems⁵¹ in primary and secondary care, and four others focused on the use of PROMs feedback in specialist mental health settings.^{41,45,47,50} Four reviews focused on use of PROMs in

oncology settings,^{42,43,46,55} one review focused on use of PROMs as a means of screening for cancer-related distress⁵⁶ and two reviews focused on the use of PROMs feedback in palliative care settings.^{59,60} One review focused on feedback of PROMs data to allied health professionals.⁶¹ Three studies attempted to identify the 'barriers and facilitators' to PROMs feedback in clinical practice,^{58,59,61} two reviews adopted a theory-driven approach to the review,^{60,62} and one review combined a 'review of reviews' with existing conceptual frameworks of PROMs feedback, but focused on synthesising the quantitative evidence.⁴³

Thus, there is a large volume of literature examining the impact of individual PROMs feedback in clinical practice, and reviewers have dissected and grouped the literature in a number of different ways, for example by condition or by setting. Furthermore, even for those reviews that have focused on the same condition or setting, differences in search methods and in inclusion and exclusion criteria have resulted in these reviews including overlapping, but different, groups of studies. For example, both Chen *et al.*⁴³ and Kotronoulas *et al.*⁴² examined the impact of PROMs feedback in oncology settings, and both reviews included both RCTs and quasi-experimental studies. Chen *et al.*⁴³ identified 27 eligible studies and Kotronoulas *et al.*⁴² identified 24 eligible studies, reported in 26 papers. However, only 16 papers are common to both reviews; 10 papers appear only in Chen *et al.*⁴³ and nine papers appear only in Kotronoulas *et al.*⁴² The reviews also vary in their synthesis methodology; most adopted a narrative overview, but those with a more narrowly focused review question used a meta-analysis.^{47,51} However, although a range of synthesis methods has been used, the reviews are dominated by traditional systematic reviews of RCTs.

It is not our intention to provide a detailed analysis of the findings of each review. Here we present a brief overview of their findings in order to highlight outcome patterns that will be explored during our synthesis. Those adopting a traditional systematic review methodology to survey the entire literature have, in general, found it difficult to reach firm conclusions about the impact of PROMs feedback on the process and outcomes of patient care, largely owing to the heterogeneity of the intervention itself, and the wide range of indicators used to assess its impact.⁴⁸ There is some evidence to suggest that the purpose or function of PROMs feedback may influence its impact, with greater impact on patient outcomes when PROMs are used to monitor patient progress over time in specific disease populations, rather than as a screening tool.²⁴ One common pattern evident in these reviews is that PROMs feedback has a greater impact on clinician–patient communication, the provision of advice or counselling and the detection of problems than on patient management and subsequent patient outcomes.^{48,49}

This general conclusion is also mirrored in the reviews focusing on oncology.^{42,43} For example, Chen *et al.*⁴³ found 'strong' evidence that the feedback of PROMs data improves patient–clinician communication, and 'some' evidence that it improves the monitoring of treatment response and the detection of patients' problems. However, they found 'weak but positive evidence' that PROMs feedback leads to changes in patient management, and 'a great degree of uncertainty' regarding whether or not PROMs feedback improves patient outcomes. Chen *et al.*⁴³ suggested that greater impact of PROMs feedback may be found where PROMs are fed back for a sustained period of time to multiple stakeholders, with feedback that is clear and easy to understand, and sufficient training for health professionals. Kotronoulas *et al.*⁴² found significant increases in the frequency of discussions 'pertinent to patient outcomes', but little impact on referrals or clinical actions in response to PROMs data. This suggests that there may be a 'blockage' between the identification and discussion of the issues raised by PROMs and the ways in which clinicians respond to these issues.

The review of qualitative evidence⁵⁸ provides some further possible explanations for these findings, which can be explored in our synthesis. This review found that clinicians sometimes questioned the validity of PROMs data, and expressed concerns about the lack of clarity regarding whether PROMs data were intended for use to inform clinical care or to monitor the quality of the service. PROMs feedback was more likely to inform patient management when it provided new information to clinicians. This review also identified a number of unintended consequences of PROMs feedback. In line with some of the theories we discuss in *Chapter 8*, the intrusive nature of incorporating discussion of PROMs data into the consultation

was, in some circumstances, perceived to affect the patient–health-care practitioner interaction. They found some evidence that, rather than open up the consultation, PROMs feedback may narrow its focus, and that certain questions may distress patients and, thus, damage the patient–health-care practitioner relationship.

Thus, evaluating and reviewing the evidence of PROMs feedback is a challenge for several reasons, all of which arise from the complexity of the intervention. First, PROMs feedback is unavoidably heterogeneous and varies by PROM used, the purpose of the feedback, the patient population, the setting, the format and timing of feedback, the recipients of the information and the level of aggregation of the data.¹² Therefore, there is a need for review methods that explicitly take into account the heterogeneity of the intervention, and seek to understand how this shapes intervention success.

Second, the implementation chain from feedback to improvement has many intermediate steps and may only be as strong as its weakest link.⁶² At an individual level, PROMs feedback may improve communication and detection of patient problems, but may have less impact on patient management or health status.⁴⁸ However, its impact on communication during the consultation is not uniform and depends on the nature of patients' problems. In oncology, where there is most evidence that PROMs influence communication, clinicians were more likely to discuss symptoms with their patients in response to PROMs feedback, but not psychosocial issues.^{63,64} We are confronted with the cautionary hypothesis that PROMs feedback may not result in further discussion or the offer of symptomatic treatment because high PROMs scores (suggesting high disease impact) do not always represent a problem for the patient or a problem that clinicians perceive as falling within their remit to address.⁶⁵

At an aggregate level, there are many organisational, methodological and logistical challenges to the collation, interpretation and then utilisation of PROMs data.⁶⁶ These include reducing the risk of selection bias, as older, sicker patients are less likely to complete PROMs;⁶⁷ reducing the variation in recruitment rates in PROMs data collection across NHS trusts;⁶⁸ ensuring that procedures are in place to adequately adjust for case mix;^{69,70} collecting the data at the right point in the patient's pathway; and summarising this information in a way that is interpretable to different audiences.⁷¹ In summary, a number of potential obstacles may prevent or lead to partial success in PROMs feedback achieving its intended outcome of improving patient care. There is a need to pinpoint these obstacles or blockages more systematically in terms of their location in the implementation chain, and to identify the circumstances in which they occur and those in which they can be overcome.

Third, the success of PROMs feedback is context dependent, and these contextual differences influence the precise mechanisms through which it works and, thus, its impact on patient care. For example, using PROMs data as an indicator of service quality for surgical interventions in acute care is very different from their use as a quality indicator of GPs' management of long-term conditions in primary care. The impact of surgery on disease-specific PROMs and knowledge of the natural variability of scores has been well documented,⁷² but this knowledge is lacking regarding the impact of primary care on PROM scores.⁷³ At an individual level, surgeons are specialised and need only interpret the PROMs data in their specialty. In contrast, GPs manage patients with different long-term conditions, and need to make sense of data from different PROMs, or to disentangle the impact of different conditions on PROMs scores. The interpretation of the meaning of changes is, therefore, very different in each context.

Furthermore, differences in context can result in the intervention not working through the intended mechanisms, leading to unintended consequences.⁷⁴ For example, the feedback and public release of performance data may stimulate improvement activity at hospital level through increased the involvement of leadership or a refocusing of organisational priorities,⁷⁵ but it has also been shown to lower morale, and may focus attention on what is measured to the exclusion of other areas.¹⁸ Others have cautioned that it may also lead to surgeons refusing to treat the sickest patients to avoid poor outcomes and lower publicly reported ratings.⁷⁴ Data from the national PROMs programme have been misinterpreted by some as indicating that a significant proportion of varicose vein, hernia and hip and knee replacement should not take place.⁷⁶ Public reporting of performance data may not improve patient care, as intended, through

informing patient choice.^{19,77} Rather, patients are often ambivalent about performance data and rely on their GP's opinion when choosing a hospital.^{78,79} Thus, there is need to highlight the potential unintended consequences of PROMs feedback and to distinguish between the circumstances in which they arise.

Fourth, PROMs have been implemented against a backdrop of other initiatives designed to drive up the quality of patient care, which can potentially either support or derail the intended impact of PROMs feedback. For example, Quality and Outcomes Framework (QOF) payments are dependent on the use of a standardised questionnaire for depression screening, resulting in GPs sometimes avoiding coding a person as suffering from depression in order to circumvent the completion of a questionnaire viewed by many GPs as unnecessary.^{80,81}

Finally, despite PROMs feedback having many functions and aspirations, research coverage of them is uneven, with more studies (trials and qualitative case studies) examining PROMs feedback at an individual level and few studies examining their use as a performance indicator at a group level.

Aims and objectives

The purpose of this review is to take stock of the evidence to understand by what means and in what circumstances the feedback of PROMs data leads to the intended service improvements. For any application of PROMs feedback, its impact on the quality of patient care depends on a long, complex chain of inputs and outputs, and is greatly affected by where and how it is implemented. This complexity has made it difficult for existing systematic reviews to provide a definitive answer regarding whether PROMs feedback leads to improvements in patient care at either the individual patient level⁴⁸ or the level of health-care organisations.¹⁹ In this project, we will use a different review method, realist synthesis,⁸² to clarify how the different applications of PROMs feedback are intended to work and to identify the circumstances under which PROMs feedback works best and why, in order to inform its future implementation in the NHS.

As the applications of PROMs data continue to multiply, our first aim is to identify and classify the various ambitions of PROMs feedback. At the individual level, PROMs data are utilised to improve patient care by (1) screening for undetected problems, (2) monitoring patients' problems over time and (3) involving the patient in decisions about their care. At the group level, PROMs data may improve patient care by (4) improving the appropriateness of the use of interventions, (5) stimulating QI activities through benchmarking provider performance or (6) informing decision-making about choice of provider.⁶ Our objectives are to:

- Produce a comprehensive taxonomy of the 'programme theories' underlying these different functions, and capture their subtle differences and the tensions that may lie between them.
- Produce a logic model of the organisational logistics, social processes and decision-making sequences that underlie the collation, interpretation and utilisation of PROMs data. We will use this model to identify the potential blockages and unintended consequences of PROMs feedback that may prevent the intervention from achieving its intended outcome of improving patient care. This will provide a framework for the review.

To inform the future implementation of PROMs feedback, our second aim is to test and refine these programme theories about how PROMs feedback is supposed to work against existing evidence of how it works in practice. We will synthesise existing evidence on each application of PROMs feedback including, where necessary, evidence from other quality reporting initiatives. The specific objectives of this synthesis are to:

- identify the implementation processes that support or constrain the successful collation, interpretation and utilisation of PROMs data

- identify the mechanisms and circumstances through which the unintended consequences of PROMs data arise and those through which they can be avoided.

Our third aim is to use the findings from this synthesis to identify what support is needed to optimise the impact of PROMs feedback and distinguish the conditions (e.g. settings, patient populations, nature and format of feedback) in which PROMs feedback might work best. We will produce guidance to enable NHS decision-makers to tailor the collection and utilisation of PROMs data to local circumstances and maximise its impact on the quality of patient care.

During initial project team discussions, we established that the feedback and public reporting of aggregate PROMs data to stimulate QI efforts by providers was based on a different set of programme theories from the feedback of individual PROMs data in the care of individual patients. Therefore, to meet these aims and objectives, we decided to carry out two separate, albeit related, reviews:

- Review 1, which examined the feedback and public reporting of aggregate PROMs data to providers, aimed to explore in what circumstances and through what processes the feedback of *aggregate* PROMs data leads to improvements in patient care.
- Review 2, which examined the feedback of individual PROMs data to clinicians, aimed to explore in what circumstances and through what processes the feedback of *individual* PROMs data leads to improvements in patient care.

Public and patient involvement

We involved patients in a number of ways throughout the reviews. Laurence Wood, a public and patient involvement (PPI) representative, was a member of the project team throughout the review. He attended project team meetings, helped to inform the development of our programme theories and commented on our findings. He chaired a PPI group consisting of three PPI members, Gill Riley, Eileen Exeter and Rosie Hassaman. The group met twice during the project; this was less often than we had anticipated and was a result of a long-term condition of one of the members. The group helped to inform our programme theories and reviewed our findings to date. Laurence Wood and Eileen Exeter also attended our Stakeholder Group meeting (described in *Chapters 2 and 6*) to help to focus the review. Laurence also read and commented on our plain English summary.

Rationale and overview of methodology

In this section we explain why we chose realist synthesis to conduct our review, and provide an overview of the methodology of realist synthesis. We provide a more detailed description of the application of this methodology for review 1 in *Chapter 2* and for review 2 in *Chapter 6*.

Why realist synthesis?

Realist synthesis⁸² is designed to disentangle the heterogeneity and complexity of the intervention, and to make sense of the various contingencies, blockages and unintended consequences that may influence its success. The methodology was developed by one of the coauthors of this report (RP). It is an approach that is finding increasing use in the health-care field, and a number of current and recently completed Health Services and Delivery Research projects are making use of the approach (e.g. project 11/1022/04 led by Pawson with Greenhalgh as coapplicant, project 13/97/24 led by Wong and project 14/194/20 led by Burton). Pawson was also a team member of another key Health Services and Delivery Research project, 10/101/51; 'Realist And Meta-narrative Evidence Synthesis: Evolving Standards – RAMESES', which led to the development of reporting standards for realist synthesis.⁸³ The methodology now forms one of the

approaches used by the National Institute for Health and Care Excellence (NICE) to develop public health guidance.⁸⁴

We have chosen to use realist synthesis because it:

- permits us to understand in what circumstances and through what processes the feedback of PROMs data improves patient care and why (rather than just answering 'does it work?')
- recognises that the success of PROMs feedback is shaped by the ways in which it is implemented and the contexts in which it is implemented
- allows us to combine evidence from different types of empirical studies (both qualitative and quantitative).

What is realist synthesis?

We do not intend to provide a detailed description of the origins and basic assumptions underpinning realist synthesis; this can be found elsewhere.⁸² However, we assume that some readers of this report will not be familiar with the methodology and, therefore, we provide a basic introduction to its *modus operandi*. Realist synthesis is a review methodology that is based on the premise that social programmes or interventions constitute ideas and assumptions, or *theories*, about how and why they are supposed to work. As Pawson and Tilley argue, social programmes are theories incarnate.⁸⁵ As such, the unit of analysis of realist synthesis is not the intervention per se but the *programme theories* that underpin them. Therefore, the task of realist synthesis is an iterative process of identifying, testing and refining these programme theories to build explanations about how, and in what circumstances, these interventions work and why. In practical terms, this means that we can draw on evidence from interventions that share the same programme theories within the synthesis. For example, as we discuss in subsequent chapters of this report, the feedback of aggregated PROMs data shares many of the ideas and assumptions as the public reporting of hospital report cards and patient experience data regarding how it is intended to work. Therefore, it is legitimate to include studies that have evaluated these interventions in the synthesis; even though they are different interventions, they share the same programme theory.

Realist synthesis is also premised on the idea that it is not the intervention (in our case, PROMs feedback) itself that gives rise to its outcomes. Rather, interventions offer resources to people, and it is people choosing to act, or not to act, on these resources (known as *mechanisms*) that will determine their impact on patient care. Furthermore, complex interventions, such as PROMs feedback are never universally successful, as people differ in their response to the intervention and their responses are supported or constrained by the social, organisational and political circumstances in which PROMs feedback is implemented (*context*). What realist synthesis aims to do is *explain why* PROMs feedback works in some circumstances and not others. It does so through a process of developing, testing and refining theories about how the intervention works, expressed as context–mechanism–outcome configurations. These are hypotheses which specify that, in this situation (context), the intervention works through these processes (mechanisms) and gives rise to these outcomes.

Initially, these theories focus on practitioner, policy-maker and participant ideas and assumptions about how the intervention is intended to work (or not). These ideas can then be formulated into programme theories to specify hypotheses that certain *outcomes* (intended or unintended) will occur as a result of particular *mechanisms* being fired in particular *contexts*. As synthesis progresses and these theories are tested across a range of contexts through a review of the empirical literature, these theories are refined to develop explanations at a level of abstraction that can allow generalisation beyond a single setting. The 'end product' of realist synthesis is explanation through the formulation of 'middle-range' theories that are limited in scope, conceptual range and claims, rather than offering general laws about behaviour and structure at a societal level.⁸⁶ Middle-range theories are identified by drawing across the literature to explain why regularities in the patterns of contexts, mechanisms and outcomes occur. Thus, they provide the basis for guidance to help policy-makers to target PROMs feedback interventions to local circumstances, and highlight what support they may need to put in place in order to maximise their impact on patient care.

How is realist synthesis conducted?

Again, we do not intend to provide a detailed description of the process of conducting a realist synthesis; this can be found elsewhere.⁶² However, here we do offer a blueprint, so that readers who are not familiar with the methodology can make sense of how we operationalised our synthesis. It also makes explicit our understanding of how a realist synthesis *ought* to be conducted, which can be subject to scrutiny in judging the rigour and quality of the review.

Realist synthesis is an iterative review methodology, consisting of five main steps. For ease, these are described sequentially, but, in practice, there is considerable movement back and forth between different steps. Furthermore, a number of these steps, for example searching, quality appraisal, data abstraction and synthesis, are integrated rather than conducted separately.

Step 1: searching for and identifying programme theories

The basic unit of analysis in realist synthesis is not the intervention, but the ideas and assumptions or programme theories that underpin it. Thus, the starting point of realist synthesis is to search for and catalogue the different ideas and assumptions about how interventions are supposed to work. Initially, these programme theories focus on practitioner, policy-maker and participant ideas and assumptions about how the intervention is intended to work (or not). These may specify the sequences of steps required to deliver the intervention, and the organisational and social processes required in order for the intervention to achieve its intermediate and final outcomes: that is, an 'implementation chain'. They may also identify potential blockages in this process, as well as potential unintended consequences. They often contain ideas about the different reactions or responses that participants may have to an intervention (mechanisms) that will determine whether or not the intervention is successful (outcome). They may also include ideas about the circumstances (or context) that determine the kind of reactions participants may have to an intervention, and the blockages that may occur, which thus influence the impact of the intervention.

These ideas often remain tacit and unexpressed in empirical evaluations of the intervention, which frequently assume shared knowledge regarding how the intervention is intended to work or consider that knowing how the intervention works is not important: the task is simply to know whether or not it works. Therefore, unearthing and cataloguing these ideas is best achieved through searching and analysing policy documents, position pieces, comments, letters, editorials, critical pieces and websites or blogs that express and debate these tacit assumptions and explain how the intervention in question is intended to work. For some interventions, it can be a useful exercise to deconstruct empirical investigations or policy documents of a given intervention to surface the implicit assumptions underpinning the design of the intervention itself.⁶² This often requires a considerable amount of 'detective' work and reflection on the part of the researcher. We explain how we searched for the programme theories underlying PROMs feedback in the more detailed description of methodology in *Chapters 2 and 6*.

Step 2: focusing the review and selecting programme theories

Inevitably, the search for programme theories results in the identification of many different ideas about how the intervention is supposed to work, and its potential blockages and unintended consequences. It is not possible to review all of these, and the next stage of realist synthesis involves a process of (1) identifying common mechanisms or issues across the different programme theories, and (2) prioritising which set of programme theories to review.

The first is an important initial step in developing 'middle range' theory, which allows transferable lessons to be made. It requires the researcher to think 'what are these programme theories an example of?' and 'how do these programme theories relate to more formal or abstract theories?'. Thus, it represents a process of moving up and down a ladder of abstraction, from practitioners' ideas about how a specific intervention works, to more abstract ideas about how the family of interventions which share that programme theory are expected to work. This plays an important part in defining the boundaries of the

review, as it serves to identify other interventions that also share the same programme theory, evaluations of which might therefore be included in the theory-testing phase of the review.

The second is a process of narrowing and deciding which of these theories we might focus on. There are no set criteria to govern these decisions,⁸² but they can focus on:

- which aspects of the programme theory stakeholders and practitioners consider most important or would like to be answered
- understanding how and why one particular section of the implementation chain works or becomes 'blocked'
- considering how the same programme theory fares in different contexts
- adjudicating between rival ideas about the mechanisms through which an intervention is intended to work.

These decisions then form the framework for the review. However, it must be recognised that this process is iterative. Inevitably, the process of testing one programme theory uncovers a number of 'sub' or 'mini' theories within the review. Furthermore, the review is also likely to focus on a smaller number of main theories. Therefore, defining and redefining the boundaries of the review is an ongoing and iterative process. We explain how we identified relevant abstract theories and how we narrowed down the focus of our review in *Chapter 2* for review 1 and *Chapter 6* for review 2.

Step 3: searching for empirical evidence

The programme theories to be tested provide the backbone of the review, and determine the search strategy and decisions about study inclusion into the review in order to test and refine these theories. The next stage of the review thus involves an evidence search to identify primary studies that will provide empirical tests of each component of the theory. This involves electronic database searches, as well as forwards and backwards citation tracking. Searching and synthesis are interwoven, and, as the synthesis progresses, the emergence of new subtheories or mini theories often requires further iterative searches to identify empirical evidence to test them. Furthermore, the review is also likely to focus on a smaller number of main theories as the synthesis progresses. In *Chapters 2* and *6*, we describe in some detail the processes we used to identify the empirical evidence on which this review is based.

Step 4: quality appraisal and data extraction

These are combined in realist synthesis. Different programme theories require substantiation in divergent bodies of evidence. Hypotheses about the optimal contexts for the utilisation of PROMs data are tested by comparing the outcomes of experimental studies in different settings, claims about the reactions of different recipients of PROMs data are tested using qualitative data, etc. Studies (or parts thereof) are included in the study depending on their relevance to the programme theory being tested.

Quality appraisal is conducted throughout the review process, and goes beyond the traditional approach that focuses on only the methodological quality of studies.⁸⁷ In realist synthesis, the assessment of study rigour occurs alongside an assessment of the relevance of the study, and occurs throughout the process of synthesis. Quality appraisal is done on a case-by-case basis, as appropriate to the method utilised in the original study. Both qualitative and quantitative data are compiled. In addition, the inferences and conclusions drawn by the authors of the studies are also extracted as data in realist synthesis, as they often permit the identification of subtheories that can then be further tested with empirical evidence. Different fragments of evidence are sought and utilised from each study. Each fragment of evidence is appraised, as it is extracted, for its relevance to theory testing and the rigour with which it has been produced.⁸⁷ In many instances, only a subset of findings from each study that relate specifically to the theory being tested are included in the synthesis. Therefore, quality appraisal relates specifically to the validity of the causal claims made in these subset of findings, rather than the study as a whole. Trust in these causal claims is also enhanced by the accumulation of evidence from a number of different studies, which provides further lateral support for the theory being tested, discussed in more detail in the following section. Finally, quality appraisal is integrated into the synthesis narrative, rather than reported separately.

Step 5: synthesis

The goal of realist synthesis is to refine our understanding of how the programme works and the conditions and caveats that influence its success, rather than offering a verdict, descriptive summary or mean effect calculation on an intervention or family of programmes. Synthesis takes several forms. At its most basic, realist synthesis involves building 'lateral support' for a theory by bringing together information from different primary studies and different study types to explain why a pattern of outcomes may occur. Another form of synthesis, particularly useful when there is disagreement on the merits of an intervention, is to 'adjudicate' between the contending positions. This is not a matter of providing evidence to declare a certain standpoint correct and another invalid. Rather, adjudication assists in understanding the respects in which a particular programme theory holds and those in which it does not. Finally, the main form of synthesis is known as 'contingency building'. All PROMs feedback programmes make assumptions that they will work under implementation conditions A, B, C, applied in contexts P, Q, R. The purpose of the review is to refine many such hypotheses, enabling us to say that, more probably, A, C, D, E and P, Q, S are the vital ingredients. In *Chapter 2*, we will provide short examples of how we carried out our synthesis.

Structure of the report

This report is divided into two parts. Review 1, consisting of *Chapters 2–5*, reports our realist synthesis of the feedback of aggregate PROMs and performance data to providers. Review 2, consisting of *Chapters 6–9*, considers the feedback of individual PROMs data to inform the care of individual patients.

Review 1: a realist synthesis of the feedback of aggregate patient-reported outcome measures and performance data to improve patient care

Chapter 2 provides a description of the methodology of review 1, and details the process of searching for programme theories, the process of searching for evidence to test these theories, how studies were selected for inclusion in the synthesis, and how data were extracted and synthesised. In *Chapter 3*, we provide a comprehensive taxonomy of the ideas and assumptions, or programme theories, underlying the feedback of aggregate PROMs data to providers. In *Chapters 4 and 5*, we report the findings of our evidence synthesis for the feedback of PROMs and other performance data to providers. *Chapter 4* interrogates the mechanisms through which this is intended to occur, while *Chapter 5* considers how different contextual configurations influence which of these mechanisms occur and the subsequent intended (or unintended) outcomes.

Review 2: a realist synthesis of the feedback of individual patient-reported outcome measures data to improve patient care

Chapter 6 provides a description of the methodology for review 2, again detailing the process of searching for programme theories and for evidence to test these theories, and the ways in which we selected studies for inclusion and extracted and synthesised data. *Chapter 7* examines the programme theories underlying the feedback of PROMs data at the individual level, offering a taxonomy on the ideas and assumptions, or programme theories, underlying the feedback of individual PROMs data. In *Chapters 8 and 9*, we report on the evidence synthesis of the implementation chain through which the feedback of PROMs data in the care of individual patients is expected to work. In *Chapter 8*, we explore the circumstances in which, and process through which, PROMs completion may support patients to raise issues with clinicians. In *Chapter 9*, we examine clinicians' use of PROM feedback to support their care of individual patients. Finally, *Chapter 10* brings our findings together and discusses their implications for practice and future research.

Chapter 2 Review methodology: feedback and public reporting of aggregate patient-reported outcome measures data

A protocol for our realist synthesis has been published,⁸⁸ and in this chapter we describe how the boundaries and focus of our review of the feedback of aggregate PROMs data were defined. This enables the reader to understand why and how changes were made to the original protocol, as suggested by the RAMESES guidelines for reporting realist syntheses.⁸³ We describe how we carried out our review, using the five steps outlined in *Chapter 1* as a structure.

Searching for and identifying programme theories

As discussed in *Chapter 1*, identifying opinions and commentaries for a realist synthesis is the first stage in identifying theories for which evidence is later sought. The purpose of this search is to map the range and diversity of different programme theories underlying PROMs feedback, rather than identify and include every single paper discussing the ideas and assumptions underlying PROMs feedback. We conducted one search for programme theories for PROMs feedback at both the aggregate and individual level. Opinion pieces and commentaries on PROMs feedback were identified in database searches, JG's personal library (89 known relevant studies) and citation tracking activities including forwards and backwards citation searching.

In April–May 2014, we searched the following databases:

- Cochrane Database of Systematic Reviews (via Wiley Online Library), issue 5 of 12, May 2014
- Cochrane Methodology Register (via Wiley Online Library), issue 3 of 4, July 2012
- Database of Abstracts of Reviews of Effects (via Wiley Online Library), issue 4 of 4, October 2014
- EMBASE Classic+EMBASE (via Ovid), 1947–30 April 2014
- Health Management Information Consortium (via Ovid), 1983–present
- (Ovid) MEDLINE®, 1946–week 3 April 2014
- (Ovid) MEDLINE® In-Process & Other Non-Indexed Citations, 1966–29 April 2014
- NHS Economic Evaluation Database (via Wiley Online Library), issue 4 of 4, October 2014.

Two search strategies were run on the Ovid databases: one aimed at identifying review papers and one aimed at identifying commentaries and opinion pieces. The Cochrane Library databases were searched with one strategy to identify reviews only, as they were unlikely to contain opinion pieces. The searches were developed iteratively; initial searches developed by the information specialist (JW) were discussed with JG and SD, who provided feedback on whether or not useful papers were being captured. JW then revised the search strategy.

All search strategies included search concepts for PROMs and the 'Outcomes of Feedback'. Subject headings and free-text words were identified for use in the search concepts by JW and project team members. Further terms were identified and tested from the personal library (known relevant) papers.⁵⁷ Care was taken to avoid retrieving papers that simply reported PROMs outcomes, and to identify those with discussion of the feedback of PROMs.

An 'opinion pieces' search strategy from a previous realist synthesis,⁸⁹ conducted by the same authors, was tested against the known relevant papers and used (with a minor adaption in MEDLINE to include the search term 'comment.cm'). An example of the PROMs feedback 'opinion pieces' search is presented in *Appendix 1*. The search strategies for review papers used the Clinical Queries – Reviews specificity

maximising filter in Ovid databases plus a series of specific free-text searches to identify reviews (see *Appendix 1*).

The database searches identified 1011 references, which reduced to 837 when duplicates were removed. These records were stored in an EndNote (Thomson Reuters, CA, USA) library alongside the 89 personal library references to create set of 748 references.

JG and SD screened the titles and abstracts of the 748 references to identify potentially relevant papers according to the following criteria.

Inclusion criteria

- The paper describes how aggregate PROMs feedback is intended to work.
- The paper provides a theoretical framework that describes how aggregate PROMs feedback is intended to work.
- The paper provides a critique of the ideas underlying how aggregate PROMs feedback is intended to work.
- The paper reviews ideas about how aggregate PROMs feedback is intended to work.
- The paper provides stakeholder accounts or opinions of how aggregate PROMs feedback does/does not work.
- The paper outlines, discusses or reviews potential unintended consequences of aggregate PROMs feedback.

Exclusion criteria

- The paper reports findings in which a PROM is used as a research tool [e.g. an evaluation of an intervention, a study exploring the health-related quality of life (HRQoL) of specific populations].
- The paper is focused on evaluating the psychometric properties of a PROM.
- The paper reviews the psychometric properties of a PROM or a collection of PROMs.
- The paper provides advice or recommendations about which PROM to use in a research context.

An initial screen identified 94 potentially useful papers; the titles and abstracts of these papers were then rereviewed by JG and categorised according to the different theories they articulated. All of these papers contributed to the process of mapping the programme theories underlying PROMs feedback. Following this process, 46 were selected for inclusion, as they provided the clearest examples of the ideas underpinning the feedback and public reporting of PROMs and performance data. These papers represented the same ideas and assumptions contained within the full set of 94 papers and, in essence, were a purposive sample of these papers. The full texts of these papers were then read, together with an additional 30 papers from JG's existing library, which included key policy documents and grey literature. Notes were taken about the key ideas and assumptions regarding how the feedback and public reporting of PROMs data was intended to work. Of the 46 papers identified from the literature searches, 15 were purposively selected as 'best exemplars' of the ideas reflected in the papers as a whole. These were cited in the final draft of the paper cataloguing and summarising the programme theories underlying the feedback and public reporting of aggregated PROMs data, reported in *Chapter 3*. However, again, they represented similar ideas and assumption as the 94 papers initially selected for inclusion. Forwards and backwards citation tracking of key articles, and additional, iterative searches in Google Scholar (Google, Inc., Mountain View, CA, USA) as key subtheories emerged, identified a further 30 papers that were cited in the final draft, reported in *Chapter 3*. In all, 75 papers contributed to the development of programme theories. Regular discussion of these ideas among the project group, circulation and feedback of draft working papers outlining the theories ensured that the full range of different theories were represented. *Figure 1* summarises the flow of studies from identification through to inclusion in the final document.

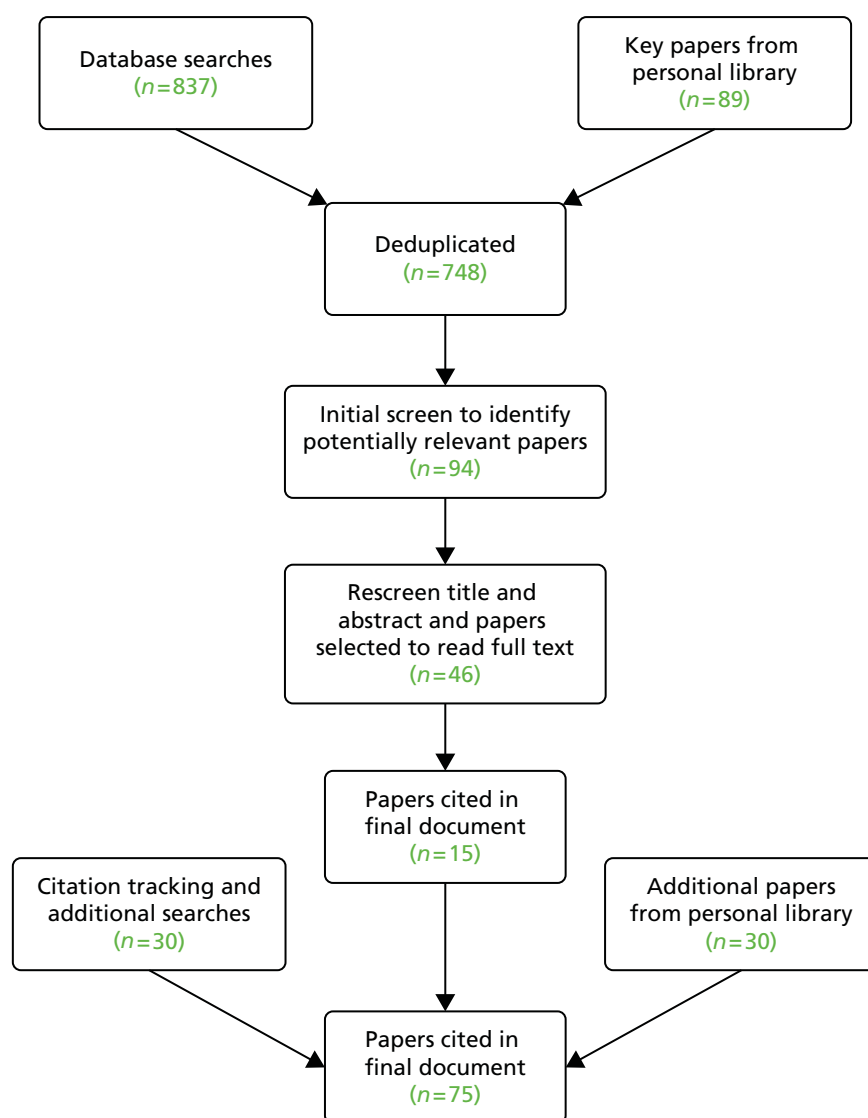


FIGURE 1 Summary of searching and selection process for aggregate programme theories.

Alongside these searches, we also held informal meetings with a number of stakeholders. These included the insight account manager for PROMs and a senior analyst from NHS England, an information analyst for the national PROMs programme from the Health and Social Care Information Centre (HSCIC), a GP commissioner from a Clinical Commissioning Group and a NHS trust lead for patient experience. In these meetings, we explored stakeholders' views on how the national PROMs programme and the national surveys of patient-reported experience measures (PREMs) were intended to improve patient care. These interviews were neither tape recorded nor formally analysed. Rather, they were used to clarify our ideas about how these programmes were intended to work, and to support and expand the programme theories we identified from the literature.

Focusing the review and selecting programme theories

Agreeing the focus of the review

The process of cataloguing the different programme theories underlying PROMs feedback at the aggregate level (reported in *Chapter 3* of this report) allowed us to identify the inner workings of these interventions as perceived by those who design, implement and receive these interventions. To agree the focus of the

review, we presented our initial programme theories and a basic logic model of the feedback of aggregate PROMs data to our patient group and at a 1-day stakeholder workshop (also attended by two members of our patient group). Our patient group consisted of three 'expert' patients: one was a retired GP, one had previously worked for a NHS Commissioning Board and the third worked for a national charity, Arthritis UK. Our stakeholder event included the following stakeholders:

- three analysts on the national PROMs programme from NHS England
- analyst working on the national PROMs programme from the HSCIC
- Matron for Surgery, Anaesthesia and Theatre
- Senior Sister for Surgical Pre-Assessment
- Director of Operations at a NHS trust
- representative from the Royal College of Nursing with expertise in PROMs
- consultant surgeon
- two academics with expertise in orthopaedics and PROMs
- two patient representatives.

We presented our initial programme theories as 'propositions about how PROMs feedback is intended to work' at these meetings, and invited participants to comment on these ideas and refine, extend and prioritise them. Stakeholder discussions focused on aspects of the national PROMs programme that they found challenging, which included:

- Variations in how providers used the PROMs data provided, which was perceived to depend on the size of the trust's information technology (IT) department; and thus the resources available to interrogate and analyse these data.
- Variations in how PROMs data were disseminated, in terms of who these data were shared with and how.
- Whether or not staff on the ground felt that PROMs data provided information to enable them to identify the causes of poor care and solutions to address them.
- Whether or not PROMs data could be linked or interpreted in relation to other locally collected data to enable trust boards to utilise the data effectively.
- There was some scepticism about whether or not PROMs data would inform patient choice of hospital; patients felt that they were more likely to be used to reassure patients that the care they received was of a high standard (i.e. for public accountability).

Following the stakeholder workshops, we held a project team meeting to reflect on the issues raised and agree the focus of the review. Although stakeholders had not actively prioritised our theories in the workshop, they had provided a valuable perspective on 'why PROMs feedback may not work as intended'. It was felt that much research had focused on how the data are collected, but the key issue that emerged from the stakeholder workshop was the difficulties that providers experienced in responding to the data. Therefore, we decided to focus our review on how different stakeholders were expected to respond to the feedback and public reporting of PROMs data. There was some debate about whether or not the review should consider the use of PROMs as a tool for patient choice. The stakeholder group had talked about PROMs as a means of public accountability in terms of providing reassurance to the public of good care, rather than informing decisions about which hospital to go to. Our patient group had expressed some scepticism about the idea that PROMs data would be used by patients to inform their choice of hospital. We also recognised that systematic reviews had found little evidence that patients used performance data to inform their choice of hospital.¹⁹ Therefore, we agreed we would focus on how providers were expected to respond to PROMs data. Thus, our review question was:

- In what circumstances and through what processes do providers respond to the feedback and public reporting of PROMs data to improve patient care, and why?

However, during the review itself, we found that the process of testing the hypothesis that providers respond to performance feedback to protect their market share led us to synthesise some aspects of the literature exploring how patients made choices about which hospital to attend.

Identifying and searching for abstract theories

The next step in our review was to make connections between these lower-level practitioner theories and higher-level, more *abstract* theories to develop a series of hypotheses that could be tested against empirical studies in our review, and produce transferable lessons about how and in what circumstances PROMs feedback produces its intended outcomes.

To identify the abstract theories, JG, SD and RP engaged in a series of joint brainstorming sessions and analysis of the PROMs programme theories. The aim of these sessions was to identify the abstract, higher-level theories relevant to PROMs feedback. To do this, we tried to answer the questions ‘what is this intervention an example of?’, ‘what is the core underlying idea at work here?’ and ‘what other interventions also share these ideas?’ We identified three key ideas underlying PROMs feedback at the aggregate level:

1. Audit and feedback: PROMs feedback is an example of an audit and feedback intervention, as it involves generating a ‘summary of clinical performance of health care over a specified period of time aimed at providing information to health professionals to allow them to assess and adjust their performance’.⁹⁰
2. Benchmarking: PROMs feedback also has a comparative element, such that providers can also compare their own performance with that of other providers in their locality or across England.
3. Public disclosure: PROMs data are made publicly available to a range of stakeholders, including patients, who are expected to exert pressure on providers to improve patient care.

We discuss these ideas in more detail in *Chapter 3*. To identify abstract theories relating to these ideas, we conducted searches using search terms ‘feedback NHS’, ‘benchmarking NHS’, ‘audit and feedback NHS’ on Google and Google Scholar (August 2014). For each search, we screened the first five pages and selected papers according to the following criteria:

- presents or discusses an abstract theory
- presents propositions about how (mechanisms) and in what circumstances (contexts) the intervention may work best
- contains a map, model or implementation chain of the theory.

From 300 references, we identified six papers as meeting our inclusion criteria. For each selected paper, we checked the references to identify other related papers and also undertook a process of forwards citation tracking of four key papers^{91–94} to identify papers that cited these. We also consulted existing systematic reviews^{90,95,96} to identify references to abstract theories.

The searches were run August 2014 in the following resources:

- Google Scholar
- Sciences Citation Index (via Thomson Reuters Web of Science) – 1900–present.

These searches identified 69 references, which were reduced to 65 when duplicates were removed. We drew on a total of 27 papers from all of these searches to inform our thinking about abstract theories underlying PROMs feedback, which formed the basis of a working paper discussed among the project team. From this working paper, we cited 13 papers in our final report, which are reported in *Chapter 3*. *Figure 2* provides a summary of these searches.

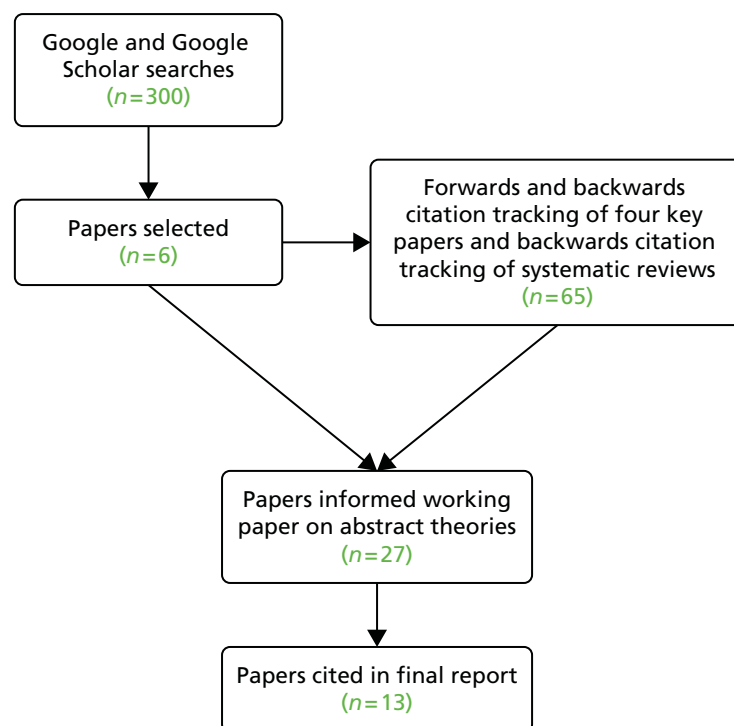


FIGURE 2 Summary of abstract theory searches.

Searching for empirical evidence and selection of studies

The next stage of our realist synthesis involved searching for empirical evidence in order to test and refine our programme theories. JW developed a search strategy, with input from SD and JG, to search for published studies against which to test our theories. We were aware that there were very few papers looking at how providers have responded to PROMs feedback per se. Following our analysis of abstract theories described in the previous section, we identified a number of interventions that shared the same underlying programme theory, so we also searched for studies that had evaluated these interventions. These included provider views on and responses to:

- feedback of patient experience data (e.g. the National Inpatient Survey, the GP Experience Survey, etc.)
- National Clinical Audits
- other forms of publicly reported 'performance data', for example mortality data or process data; many of these come from the USA, where there is a long history of public reporting.

The development of the search strategy was iterative; JW ran a strategy and sent JG and SD initial results; SD and JG provided feedback on whether or not the resulting papers were useful for theory testing. After several iterations, this resulted in the agreement of a final search strategy. All search strategies included search concepts for PROMs or other performance indicators, outcomes of feedback (e.g. decision-making, improved participation and communication) and qualitative research (see *Appendix 1*). In October 2014 we searched the following databases:

- EMBASE Classic+EMBASE (via Ovid) 1947–2014 October 17
- Health Management Information Consortium (via Ovid) 1983–present
- (Ovid) MEDLINE® 1946–week 2 October 2014
- (Ovid) MEDLINE® In-Process & Other Non-Indexed Citations, 1966–29 April 2014.

The searches identified 2080 records, which reduced to 1617 after removing duplicates found across the searches and duplicates of records that had already been identified from previous (theories) searches.

JG and SD independently reviewed the titles and abstracts of the first 160 references (approximately 10% of the total) using a broad set of inclusion and exclusion criteria, and compared and discussed our included studies to check that we were making comparable judgements. We then split the papers between us, and screened the titles and abstracts using a broad set of inclusion and exclusion criteria.

Inclusion criteria

- Studies about provider or commissioner views of, responses to, use of, or interpretation of PROMs data, national clinical audits, patient experience data, clinical outcomes (as an indicator of treatment effectiveness and, thus, performance) and mortality data (as an indicator of performance).
- Reports on the process of implementation of local PROMs/patient experience data collection for use as an indicator of service quality, reporting not just the results of the data collection but how data collection was implemented and/or how it was used.
- Reports on the process of implementation and use of local audit to improve care – how it was implemented and how people responded – not just reporting the results of the audit.

Exclusion criteria

- Articles on the development or validation of PROMs data, patient experience data, national clinical audits, mortality data, clinical outcomes.
- Articles about patient involvement in patient experience data.
- Articles reporting just the findings/analysis of audit data.
- Articles evaluating the impact of other QI programmes that did not involve some sort of feedback or public reporting of data.

At this stage we included 124 papers. After rereading the titles and abstracts, we developed a more restrictive set of inclusion and exclusion criteria that were focused on our evolving theories, as follows.

Inclusion criteria

Studies about how clinicians or managers have used or responded to or about clinicians' views of:

- national or local patient experience data collection and feedback
- mortality report cards or anything described as 'performance data'
- hospital report cards
- aggregate clinical outcome indicators or process data.

Studies contributing to testing theories about:

- the mechanisms through which feedback is intended to work (intrinsic desire to improve, peer comparison, protecting market share, protecting professional reputation)
- the contextual configurations that might influence how performance feedback works (financial incentives, credibility, 'actionability' of these data).

Exclusion criteria

Studies examining:

- views/experiences or implementation of general QI activities
- implementation of, or response to, local clinical audit or guidelines
- results/findings of audits or patient experience data.

Following the application of these criteria, we included 28 studies. We also checked the references of an existing systematic review of public reporting of performance data,²⁷ which identified a further 18 studies, and also checked the reference lists of five key papers^{18,75,97–99} on which we had conducted a preliminary

synthesis and identified 26 studies. We conducted an additional search in MEDLINE and EMBASE for feedback of patient experience data (see *Appendix 1*) and selected three studies from the 194 identified. This gave us a total of 78 references. We then checked for duplicates and rescreened the papers using the more focused set of inclusion and exclusion criteria listed above, and included 51 papers. These papers focused on providers' views and responses to performance data and indicators and patient experience data. *Table 1* provides a summary of the different sources of papers.

We read the 51 papers and began our synthesis, described in the next section (see *Data extraction, quality assessment and synthesis*). As the synthesis progressed, we further focused our synthesis on a smaller number of main theories but, at the same time, a number of 'subtheories' within this focused selection of 'main' theories were identified. This required us to revisit our original search results, further examine documents in JG's personal library and carry out additional citation tracking of key studies to further test these subtheories.

This identified an additional 30 papers that were included in the final synthesis. At the same time, we found that some of the papers we had originally included were no longer relevant to the main theories in the synthesis. Of the original 51 papers identified from the initial searches, 28 made it into the final synthesis; 23 papers were left out of the final synthesis because they were not relevant to the final theory testing phase of the synthesis or did not progress the theory testing further. Thus, in total, 58 papers were cited in the final synthesis of the feedback of aggregated PROMs and performance data, reported in *Chapters 4* and *5*. A flow chart of the search process is shown in *Figure 3*.

Data extraction, quality assessment and synthesis

This was an iterative process undertaken by JG, SD and EG, with feedback from the wider project group (NB, CV, DM, LW, JJ and LL). Data extraction, quality assessment, literature searching and synthesis

TABLE 1 Summary of initial searches and included studies for aggregate PROMs feedback evidence review

Source	Number identified	How screened	Number potentially relevant
A: electronic database search	1617 (after removal of duplicates)	JG and SD coscreened approximately 160 to check the application of the criteria, and then each independently screened 811 and 806	124; 28 after second screening
B: five index papers backwards citation tracking		JG went through five index papers ^{18,75,97-99} and identified potentially relevant papers	26
C: EG's personal library: patient experience measures	7	EG selected the papers and JG and SD screened them together to identify if they focused on provider responses and the use of patient experience data	3
D: citation tracking from Totten <i>et al.</i> 's ²⁷ Agency for Healthcare Research and Quality review of public reporting	40	JG went through the report and judged that potentially relevant references were those related to answering Q3 (did feedback results in changes) ($n = 8$) and qualitative studies reporting provider awareness or views of performance measures ($n = 10$) or their reported use of them in practice ($n = 22$); these were then screened to evaluate potential relevance to our review	18
E: JG patient experience searches	194 (after removal of duplicates from EMBASE/MEDLINE)	JG screened to check if papers were about provider responses to PREMs data	3
Total (after removal of duplicates)			78; 51 after second screening

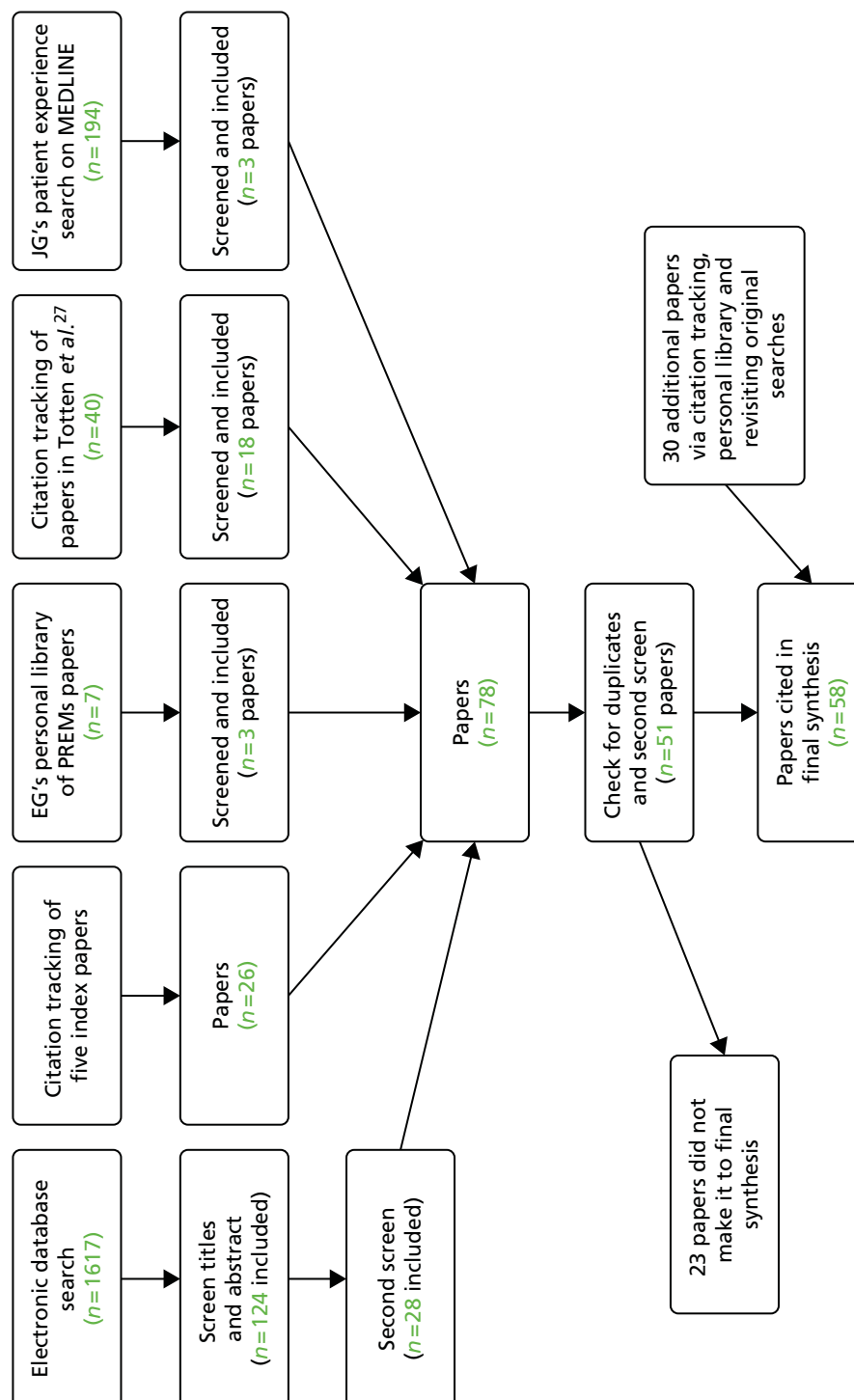


FIGURE 3 Flow diagram of the evidence search process for the feedback of aggregate PROMs data.

occurred simultaneously. To begin the synthesis process, we identified five key index papers^{18,75,97-99} from JG's personal library and conducted a 'mini' or 'pilot' synthesis on the papers. The papers were selected to represent a range of countries (the USA and the UK), settings (secondary care and primary care) and different types of performance data (PROMs, star ratings and mortality data). In this pilot synthesis, we attempted to understand:

- In what circumstances and through what mechanisms do providers interpret performance data, identify a solution to the problem and then implement that solution to improve that quality of patient care?
- In what circumstances and through what mechanisms do providers respond to performance feedback in a way that does not lead to the initiation of QI activities?

Together with our programme theories, we developed an initial logic model of how providers were expected to respond to performance data and used this as a framework to read the papers. We read the papers with our initial programme theories and logic model in mind and the theories acted as a lens through which to identify salient findings in the paper and relate them to the theory. We used the theories to make sense of the findings of each paper in and of itself, but also in comparison with other papers. After a first reading of the papers, we began to chart the potential contextual factors that might influence how performance data are responded to and the different provider responses identified within the paper. We then reread the papers to begin to develop ideas about how these different factors might come together as context–mechanism–outcome configurations: that is, ideas or hypotheses that would explain in what circumstances and through what processes performance data feedback led to initiation of QI initiatives (or not). This was both a within-paper and a cross-paper analysis. Within papers, the analysis consisted of identifying patterns in which particular clusters of contextual factors reported in a paper gave rise to a particular provider response or responses. For the cross-paper analysis, we compared these patterns across papers and attempted to explain or hypothesise why similar or different patterns might have arisen.

We discussed the findings of this pilot synthesis with the wider project group, and this informed the screening process for the papers identified via database searches and citation tracking. We screened the papers as discussed above and had an initial selection of 51 papers to begin the synthesis 'proper'. We developed a data extraction template to extract details of study title, aims, methodology and quality assessment, main findings and links to theory (see *Appendix 2* for an example of a completed data extraction template). In realist synthesis, data include not just the findings of the study but also the authors' interpretations of their findings, and we made these distinctions clear during data abstraction.

We extracted these data for all 51 studies initially, and linked the study findings to our initial programme theories. This enabled us to have a detailed understanding of each study and to begin the process of using each individual study to test and refine our theories. We discussed findings through face-to-face meetings and via Skype (Skype™, Microsoft Corporation, Redmond, WA, USA). To enable us to compare study findings in terms of the theories about when, how and why feedback and public reporting of performance data prompted providers to respond, we produced a summary table that listed key theories for each study. We considered providers in the broadest sense to mean both individuals and organisations and both primary and secondary care. However, we also recognised that different theories related to different levels of the organisations; for example, 'intrinsic' motivation theories were more relevant to individuals, whereas 'market share' theories could refer both to individuals and organisations. We then produced another table that identified common 'theory themes' across the papers (see *Appendix 3*). These included:

- whether or not market competition is necessary for providers to respond to performance data
- whether or not rewards and incentives are necessary for providers to respond to performance data
- how the quality of data collection and analysis (e.g. timing of data collection, process of case-mix adjustment) influences the extent to which it is trusted by providers
- how the 'sponsorship' of performance data influences the nature of the data collected and the extent to which they are trusted by providers
- whether low or high performers viewed or responded differently to performance data

- different intended and unintended consequences and what was perceived to be driving them (e.g. when does focusing leadership attention on poor areas of care become 'tunnel vision' and when does it gain acceptance from clinical staff and lead to change?)
- the role of the media in driving or prompting providers to respond to performance data
- whether providers are motivated to change to protect their market share because they want to be as good as or better than their peers, or from an intrinsic desire to improve care
- the relative impact of 'external' indicators and providers' own 'internal data' in driving a response to performance indicators
- the role of previous experience in public reporting, QI initiatives and involvement in broader QI activities in influencing how providers respond to public reporting
- the 'actionability' of the indicators and what makes them actionable or not.

We also produced a table summarising study findings by type of performance data, to clarify the ways in which provider responses were influenced by the characteristics of each indicator, who mandated its collection, what sanctions or incentives were attached to it and how actionable the indicator was perceived to be. We then tried to establish how these different subtheories (listed above) linked together to explain the process through which providers responded to performance data. To do this, we returned to our original logic model of the process through which providers are expected to respond to performance data. We used a whiteboard to plot the mechanisms and contextual factors that influenced the different stages of providers' responses to performance data, based on the findings of the summaries we had read. A photograph of this is shown below in *Figure 4*.

Throughout our analysis, we revised the original logic model developed in our pilot synthesis, which set out the process through which providers respond to performance data to incorporate various intermediate actions or steps that providers might take in their response, and the feedback between these two steps. We also produced a corresponding sequence of the different interlinked subtheories. *Figure 5* presents both our original and one of our revised logic models side by side (indicating how providers might be expected to respond following feedback of 'poor' performance). In the revised model, the dashed lines represent routes to 'unintended consequences', while the solid lines represent 'intended consequences'.

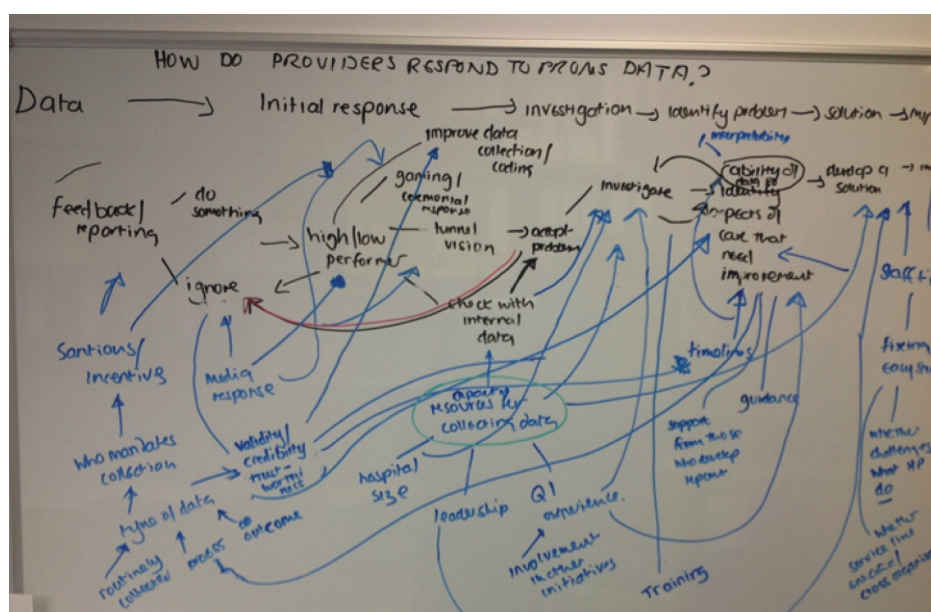


FIGURE 4 Mind map of the process through which providers respond to performance data.

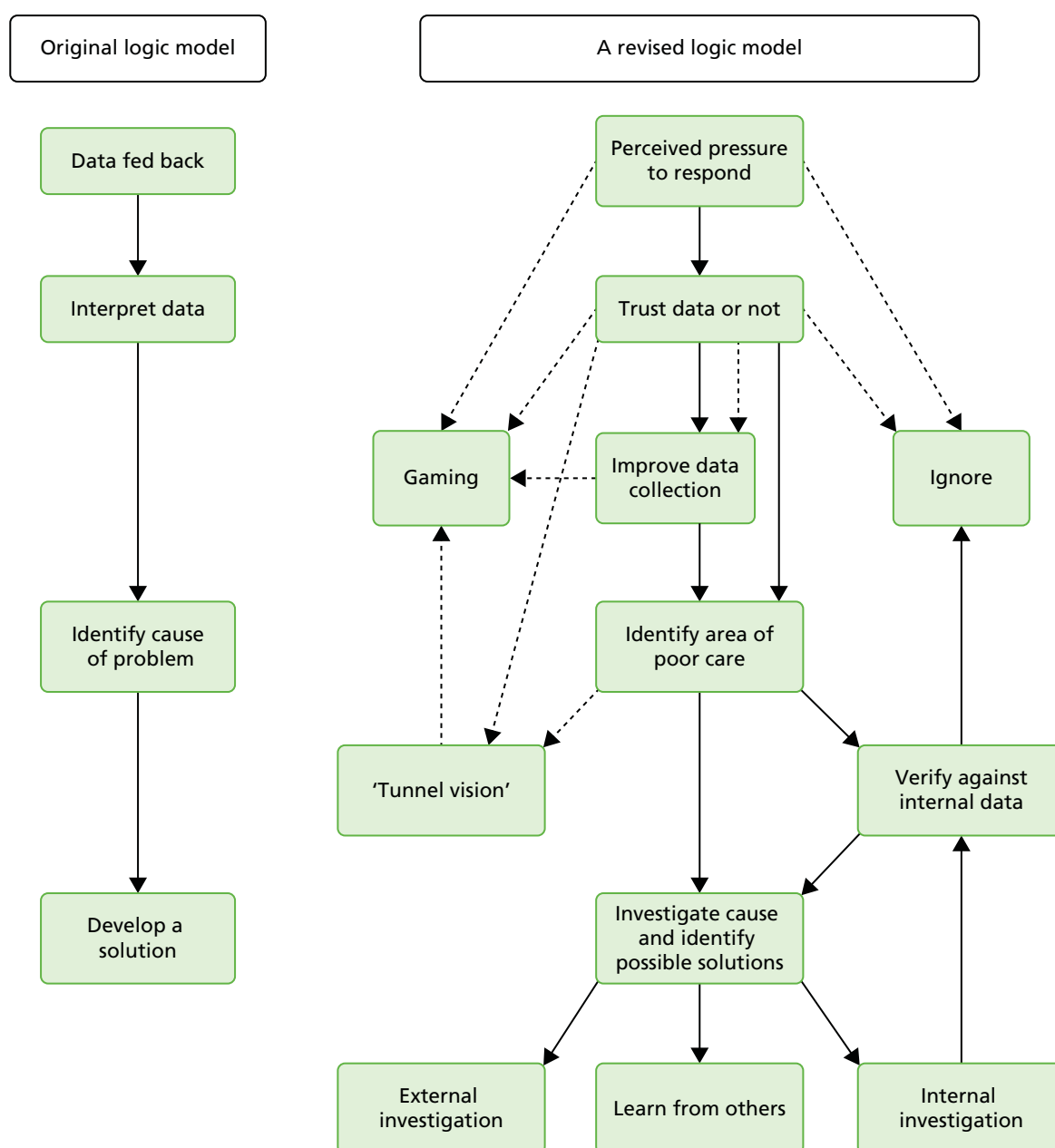


FIGURE 5 Original and a revised logic model of provider responses to performance data following feedback of 'poor' performance.

Figure 6 takes the different steps depicted in the logic model and links them to one or more of the different subtheories that seek to explain how, why and when providers may respond in a particular way. This model was revised throughout the synthesis.

As our synthesis progressed, we focused on a number of specific theories that addressed the mechanisms through which providers were expected to respond to performance data, and the key contextual factors that influenced this response. The mechanisms we explored were that providers respond to performance data because of:

- *'Intrinsic motivation' theory* Their professional ethos means that they are intrinsically motivated to maintain good patient care, and will take steps to improve if feedback highlights that there is a gap between their performance and expected standards of patient care.

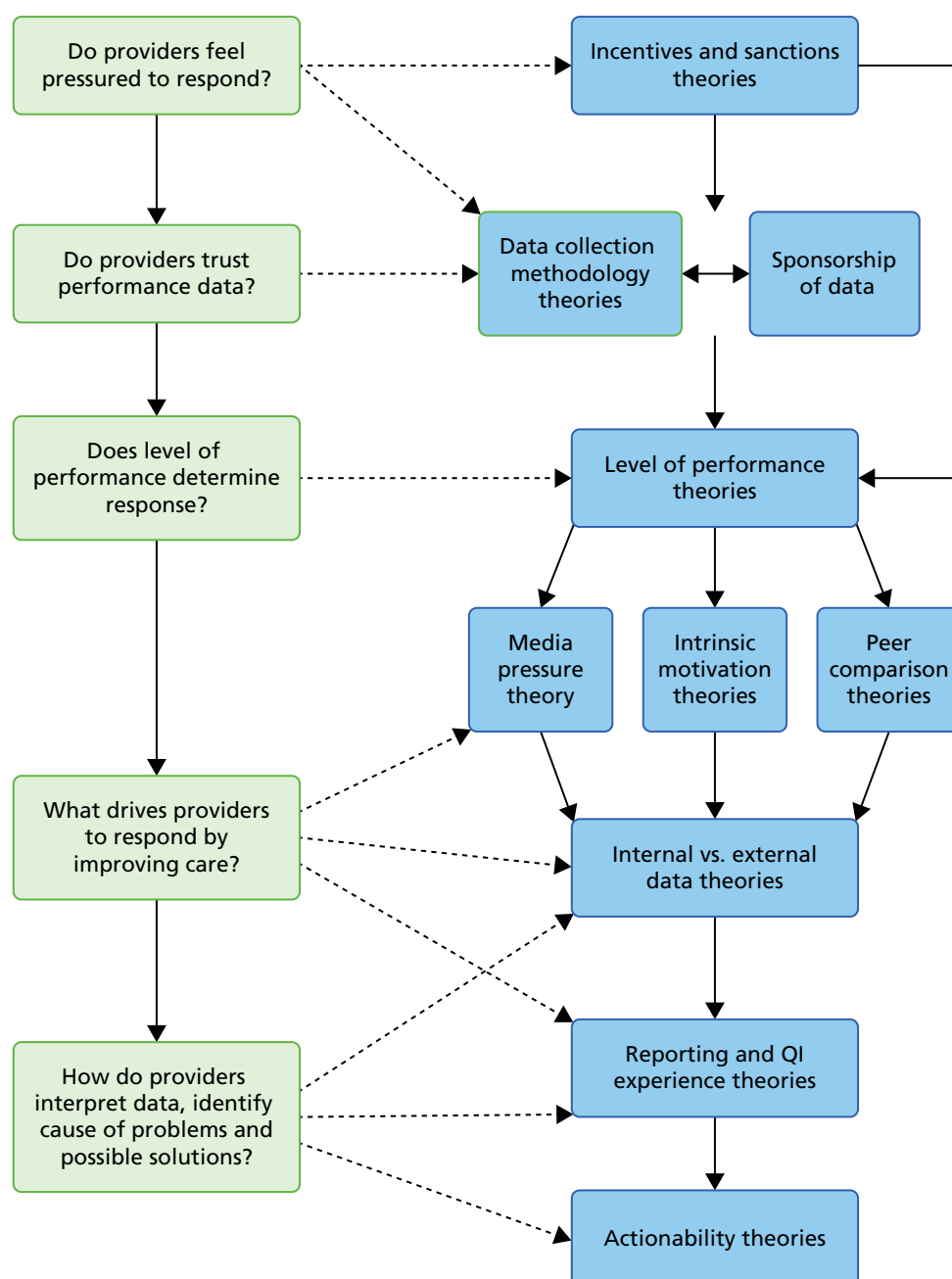


FIGURE 6 Sequence of 'middle range' theories about how and why providers respond to performance data.

- *'Market share' theory* They feel threatened by the potential loss of market share that could occur if patients decided to choose alternative, higher-performing providers.
- *'Professional reputation' theory* They wish to protect their professional or institutional reputation, which may have been damaged by being labelled a poor performer in public.
- *'Competitive benchmarking' theory* They are competitive and wish to be as good as or better than their peers.
- *'Collaborative benchmarking' theory* They improve patient care through learning about and implementing the best practices of 'high-performing' organisations as a result of the sharing of information.

The contextual factors we explored were:

- whether any rewards or sanctions were attached to performance
- the perceived credibility and validity of performance data and what determines this
- the 'actionability' of performance data.

We grouped studies together according to the theory they related to, which enabled us to test these theories by comparing and contrasting the studies, using studies to provide lateral support for a theory or, where the findings of studies differed, explaining why the studies differed. We wrote our synthesis up as a narrative account of each study and did so for a number of reasons. First, it enabled us to show how each study contributed to the theory testing process and in effect 'show our working out', so that the reader can clearly see how we came to our conclusions. Second, it enables the reader to understand how the study findings, the authors' interpretations of their findings and our interpretation of the study have contributed to the synthesis. Third, it enabled us to incorporate an assessment of the study's quality and highlight any caveats that the reader needs to be aware of in the actual narrative of the synthesis, rather than as an assessment that remains separate to the synthesis findings. We provide a number of summary sections in this narrative to enable the reader to 'take stock' of our findings, and also provide a final narrative summary of our synthesis at the end of each chapter. A summary of the theories tested in the review and included studies can be found in *Appendix 4*.

Chapter summary

In this chapter, we have described the iterative process through which we conducted our synthesis of the feedback of aggregate PROMs data. In doing so, we have followed the RAMESES guidelines⁸³ to make the process as transparent as possible. We have explained why we chose realist synthesis as our review methodology, how we searched for programme theories, how we selected our programme theories, how we searched for and selected papers, and how we synthesised the papers. In *Chapters 3–5*, we report the findings of our synthesis.

Chapter 3 Feedback of aggregate patient-reported outcome measures data: programme theory elicitation

Introduction

There is a wide range of 'big ideas' underlying how, why and in what circumstances the feedback of aggregate-level PROMs data to stakeholders will be successful in improving patient care. PROMs data are but one form of data on hospital performance, and other interventions, such as hospital star ratings in England and cardiac mortality report cards in the USA, share the same programme theories as PROMs feedback. In realist synthesis, the focus of the review is on the programme theories rather than on the intervention itself. The purpose of this chapter is to provide a thorough review and a detailed description of these intrinsic programme ideas. Note that the chapter travels no further than an exposition of these key concepts and conjectures; this is an ideas exercise. Furthermore, our aim here is not to assess the veracity of the claims made but to note their range and diversity. The assembled conjectures will go on to provide the foundation stone for further analytic work; they act as hypotheses to be tested. The basic objective of the upcoming phase of realist synthesis and subsequent chapters (see *Chapters 5, 6, 8 and 9*) is to trawl the empirical evidence to gauge how, where, why and to what extent each of these theories has proved fruitful in practice. But that is for later.

The feedback of PROMs and performance data is implemented alongside many different policy initiatives and programmes designed to improve the quality of patient care. These policies and programmes both inform the development of PROMs feedback and form the background context, which shapes the ways in which PROMs feedback works. Therefore, we begin our review by charting the policy history of quality indicators in England, including the national PROMs programme and other policy initiatives designed to improve patient care.

Finally, we conclude our chapter by making connections between these lower-level *programme* theories and higher-level, more *abstract* theories, in order to develop a series of hypotheses that can be tested against empirical studies in our review. Bringing together the programme theories and abstract theories allows us to develop a series of more general hypotheses that can be tested against the evidence, and that help produce transferable lessons about how, and in what circumstances, PROMs feedback produces its intended outcomes.

Policy and programme history

In this section, we provide a short history of the policies relating to indicators of provider performance in England, and provide concise details of the specific initiatives to arise from these. We do this for two reasons. The first is to give some understanding of how the England national PROMs programme has emerged from previous programmes and policy initiatives, and how the use of PROMs as performance indicators in other countries has risen in importance. The second is to outline the context into which PROMs feedback has been inserted, which will then be interrogated as part of our subsequent synthesis.

History of performance measurement and reporting in the NHS

The Nuffield Trust,⁸ Northcott and Lewellyn⁹² and Smith¹⁰⁰ provide useful histories of the use of performance indicators in the English NHS, beginning with those introduced by the Labour Government in 1997 in their first White Paper, *The New NHS: Modern, Dependable*.¹⁰¹ The first system to emerge was the NHS Performance Assessment Framework in 1999. Under the Performance Assessment Framework, each hospital was measured against 60 indicators, which covered health improvement, fair access, effective delivery of

appropriate health care, efficiency, patient/carer experience and health outcomes, to produce a 'balanced scorecard'. The indicators were not made publicly available, and their purpose was to assess performance across different trusts and locations in order to stimulate improvements. However, Smith¹⁰⁰ notes that the Performance Assessment Framework was used largely as a device for central government to monitor the performance of the NHS as a whole, rather than as a tool to highlight variations in performance.

In 2000, the NHS Plan¹⁰² was published, which promised increased government spending on the NHS in exchange for both structural reform and increased regulation and monitoring of NHS performance. The latter was realised through the introduction of hospital star ratings alongside a new hospital regulator, the Commission for Health Improvement (CHI). The star ratings were designed to produce a summary of hospital performance based on 40 performance indicators. Hospitals achieving three stars were judged to have the highest level of performance, while two stars were performing well overall but not consistently in every area, one-star hospitals were a cause for concern, and zero-star hospitals had the lowest level of performance against government targets. This overall rating was based on a report compiled by the CHI, following an 'Ofsted-style' inspection of the hospital alongside an analysis of performance against a set of indicators. The CHI report was fed back to trusts internally, while star ratings were made publicly available.

This system was also accompanied by a combination of financial rewards and sanctions. Trusts achieving a three-star rating were granted 'earned autonomy' in the form of less frequent monitoring and inspections for the CHI, the retention of profits from the sale of hospital land to reinvest in services, and the right to become a foundation trust. Their ratings also determined the level of discretion chief executives had to make use of the 'NHS Performance Fund' to incentivise QI at a local level. Trusts with a zero-star rating were required to produce a 'Performance Action Plan' indicating the steps taken to improve care, which had to be agreed with the Modernisation Agency and the trust's Department of Health regional office.

In 2004, the CHI was abolished and responsibility for the hospital star ratings was transferred to the newly formed Healthcare Commission. The hospital star ratings continued to be published, albeit with some modifications to the methodology through which they were produced, until 2004–5. In 2005, the Healthcare Commission introduced a new system to replace the star ratings called the Annual Healthcheck. Under this system, trust performance was graded against core standards covering seven domains of care and existing targets 'fully met', 'partially met' or 'not met'. These gradings were based on a mix of the trust's own assessment of their performance, and selective, unannounced inspections by the Healthcare Commission. Thus, responsibility for assessing performance was shifted to trusts themselves, with an external regulator acting as a 'safety net'. Trust performance on the Annual Healthcheck was published on the Healthcare Commission's website. The Annual Healthcheck continued to be published until 2008–9 and the Healthcare Commission was abolished in 2009. It was replaced by the Care Quality Commission (CQC) under the 2008 Health and Social Care Act,¹⁰³ with responsibility for regulating health and social care providers. All providers of health and social care were expected to be registered with the CQC, based on self-declaration of compliance against 16 standards. Once registered, the CQC was expected to monitor compliance with these 16 standards using routinely collected data and inspections.

In September 2013, a new model of hospital regulation was introduced. Under this model, the CQC is responsible for giving each hospital an 'Ofsted-style' rating, ranging from 'inadequate' or 'requires improvement' to 'good' or 'outstanding'. To arrive at these ratings, the CQC engages in 'Intelligent Monitoring', which involves the analysis of routinely collected data relating to hospital safety, effectiveness and patient experience, alongside informal feedback from patient organisations such as HealthWatch. PROMs data form part of this intelligent monitoring process. These data inform decisions about when and where the CQC decides to inspect hospitals and what it focuses on during visits. The CQC grading of each hospital is made publicly available on the CQC's website. Thus, regulation is intertwined with the process of judging hospital quality.

Outcomes as indicators of NHS performance

The late 2000s also saw a shift away from the measurement of structure and process as indicators of NHS performance towards a focus on outcomes. Furthermore, this has encompassed not just clinical outcomes, but also patients' assessments of their health, as measured by PROMs. In the 2008 Darzi review, *High Quality Care for All*,⁵ PROMs and PREMs were enshrined as vehicles for ensuring that patients' views of their health and experience constituted an indicator of the quality of NHS care. The importance of incorporating patients' views into the assessment of the quality of NHS care was further reinforced in the 2010 government White Paper,⁶ which promised that 'Success will be measured . . . against results that really matter to patients' (© Crown copyright; contains public sector information licensed under the Open Government Licence v3.0).

To support this focus on outcomes, the 2010 coalition government set out the NHS Outcomes Framework.¹⁰⁴ This aimed to provide a national framework through which to measure and monitor the performance of the NHS to improve patient care and make the NHS Commissioning Board accountable for the performance of the NHS. The NHS Outcomes Framework defines quality along three dimensions of (1) effectiveness, expressed in three domains; (2) patient experience; and (3) safety (Figure 7). Each domain is accompanied by a small number of overarching indicators that are agreed each year. PROMs data can be drawn on to provide indicators in domain 2. To support the monitoring of domain 3, referring to the patient experience element of the NHS Outcomes Framework, a range of new large-scale national patient surveys was introduced. These include the national GP patient survey and patient experience surveys of people attending accident and emergency, inpatients, mental health services, maternity services, and so on.

The rise of consumer choice and information

This period also saw increasing emphasis on patient choice and access to information on the quality of patient care. This move has been driven by government policy, but is also a response to public inquiries into high-profile incidences of poor care such as children's heart surgery at Bristol Royal Infirmary between 1984 and 1995¹⁰⁵ and the Francis Inquiry⁷ into Mid Staffordshire Hospital, both of which called for patients to have increasing access to information on hospital trust performance. In terms of patient choice, patients were afforded the right to choose where to have planned elective procedures and, later, to choose their own GP.⁷⁹ To support this process, 'Choose and Book' was launched in 2006, under which GPs were expected to discuss different referral options with patients, and patients could then telephone a central booking line and select a location and time of their choice for treatment. The White Paper *Equality and Excellence: Liberating the NHS*⁶ envisaged an 'information revolution', whereby patients would have much greater access to information on the quality of NHS services to inform these choices. The NHS Choices website (www.nhs.uk) was set up in 2007 to offer health and lifestyle information, advice and support to



FIGURE 7 NHS Outcomes Framework. © Crown copyright. Reproduced from Department of Health.¹⁰⁴ Contains public sector information licensed under the Open Government Licence v3.0 (www.nationalarchives.gov.uk/doc/open-government-licence/version/3/).

patients. The website contains comparative quality information on hospitals and general practices, and also includes other links to the CQC and HSCIC websites. *Table 2* summarises the information provided by the NHS Choices website for each hospital and its sources.

In addition, from April 2011, the Department of Health required all NHS hospital providers to publish 'Quality Accounts', annual reports summarising the quality of the services they deliver, which are fed back to the Secretary of State for Health and published on the trust's website and the NHS Choices website. Recently, NHS England, together with the Department of Health, the Health and Social Care Information

TABLE 2 Indicators and source of data provided for each hospital on the NHS Choices website

Indicator	Description	Source
NHS user rating	A star rating expressed from one (worst) to five (best)	NHS service users' rating of the hospital, collected via the NHS Choices website
CQC inspection ratings	A colour-coded traffic light and verbal description reflecting the CQC ratings: green star for 'outstanding', green circle for 'good', amber circle for 'requires improvement' and red circle for 'inadequate'	CQC inspection rating. A link to the CQC website is provided
Recommended by staff	Percentage of staff who agreed that if a friend or relative needed treatment they would be happy with the standard of care provided by the trust. Expressed as whether the organisation is performing as expected on this indicator (OK), worse than average or better than average	NHS Staff Survey
Open and honest reporting	Combines several other indicators to give an overall picture of whether or not the hospital has a good patient safety incident-reporting culture. Constructed from the patient safety incident reporting and response indicators used by the CQC as part of their Intelligent Monitoring system	CQC patient safety indicators
Infection control and cleanliness	Combined indicator that describes how well the organisation is performing on preventing infections and cleaning. It is constructed from the existing data displayed on NHS Choices regarding the number of <i>C. difficile</i> and MRSA infections and patients' views on the cleanliness of the ward. Overall rating for preventing infection and cleanliness: good (green), OK (blue) or poor (red)	Routinely collected surveillance data on the number of <i>C. difficile</i> and MRSA infections compiled by Public Health England; the NHS inpatient survey score for cleanliness of wards out of 10, and a rating of whether the trust that runs each hospital's score is average, below or above what would be expected; and on data from the PLACE regarding the cleanliness of the care environment, indicating if a hospital's score is in the bottom 25%, middle, or top 25% of organisations in the country
Mortality rate	The adjusted mortality ratio: deaths in hospital and within 30 days of discharge. This shows the overall rate of deaths within the NHS trust that each hospital belongs to. This includes deaths that happen while a patient is admitted to hospital and deaths up to 30 days after discharge. This mortality ratio indicator categorises NHS trusts into the bands 'better than expected', 'worse than expected' or 'as expected'	Data collated by the HSCIC
Food choice and quality	This indicator shows the results of the 2014 Patient-Led Assessments of the Care Environment, and shows a combined score for choice and quality of food. The poor (red) category shows that the hospital was in the bottom 20% of all scores for choice and quality of food. The good (green) category shows that the hospital was in the top 20% of all scores	Patient-led assessments compiled by the HSCIC

MRSA, methicillin-resistant *Staphylococcus aureus*; PLACE, Patient-Led Assessments of the Care Environment.

Service, the CQC and Public Health England, developed the My NHS website, which can be accessed from the NHS Choices website.¹⁰⁶ This website is intended to act as a tool to enable patients, as well as commissioners, providers and professionals, to access comparative data on performance in primary care, secondary care, and health and social care. The website provides access to comparative data from a range of data sources, including the national PROMs programme, the National Patient Experience Surveys, the QOF, the National Mental Health and Learning Disability Minimum Dataset, CQC inspections and mortality data for surgery, which we discuss in more detail in *Outcomes and patient-reported outcome measures collection in other NHS services*.

The UK national patient-reported outcome measures programme

The genesis of the England PROMs programme goes back to 2004, when the Department of Health commissioned a review of the psychometric properties of available PROMs suitable for the routine collection of data before and after hip replacement, knee replacement surgery, varicose veins surgery, cataract extraction and groin hernia repair.¹⁰⁷ In 2007–8, a study was carried out to assess the feasibility of collecting routine PROMs data in NHS and Independent Sector Treatment Centres.¹⁴ This report made a number of recommendations on the logistics of collection, presenting and interpreting PROMs data. In 2008, the Department of Health announced that, from 2009, the national PROMs programme would be introduced across England, with all hospitals providing care to NHS patients required to routinely collect PROMs data for patients undergoing hip replacement surgery, knee replacement surgery, varicose vein surgery or hernia repair. Thus, we can see that considerable groundwork was undertaken to determine the technical and logistical elements of data collection, but much less consideration was given to how the resulting data would be used, by whom and for what purpose.

Under the national PROMs programme, all patients who are about to undergo surgery are asked to complete a generic PROM [EuroQol-5 Dimensions (EQ-5D)] and the relevant disease-specific PROM as appropriate for hip [Oxford Hip Score (OHS)] and knee surgery (Oxford Knee Score) and varicose vein surgery (Aberdeen Varicose Vein Questionnaire). Patients return their questionnaires to a data supplier, who scans and scores them. The percentage of patients who return questionnaires for each hospital, expressed as the number of patients who return questionnaires divided by the number of patients undergoing each procedure, is known as the participation rate. Patients are then sent a follow-up disease-specific and generic PROM (3 months post intervention for varicose vein and hernia repair patients and 6 months postoperatively for hip and knee patients). The degree of health gain (i.e. the difference between the 'before' and 'after' PROM scores) for all patients who have the procedure in each hospital is calculated and adjusted for a number of case-mix variables.¹⁰⁸ Figure 8 shows a flow chart of this process.

From 2009 to 2011, the degree of health gain for each provider was available on the HSCIC's website as a downloadable spreadsheet. Since 2011, the information has also been made available as a 'funnel plot' (Figure 9). For each procedure and for each measure, the health gain experienced by patients undergoing the procedure is plotted for each hospital on a graph, with the volume of procedures on the x-axis and the amount of health gain on the y-axis. The England average is presented as the red line and outer control limits are plotted to demarcate providers that are two (95% control limits) or three (99.8% control limits) standard deviations higher or lower than the average health gain for England.¹⁰⁹ Providers with a health gain that is greater than two but less than three standard deviations below the England average are labelled as 'alerts'; providers with a health gain that is greater than three standard deviations below the England average are labelled as 'alarms'. As of 2014, patients have also been able to access these data via the My NHS website.¹⁰⁶

Outcome and patient-reported outcome measures collection in other NHS services

The national PROMs programme is not the only example of the routine collection and public disclosure of outcome or PROMs data in the NHS. Many professional bodies have engaged in clinical audits, which often include outcome data, for a number of years. For example, the Healthcare Quality Improvement Partnership, an independent, professionally led organisation, currently supports over 30 national clinical audits. Summary data on each audit in the form of reports for different audiences are publicly available from the Healthcare

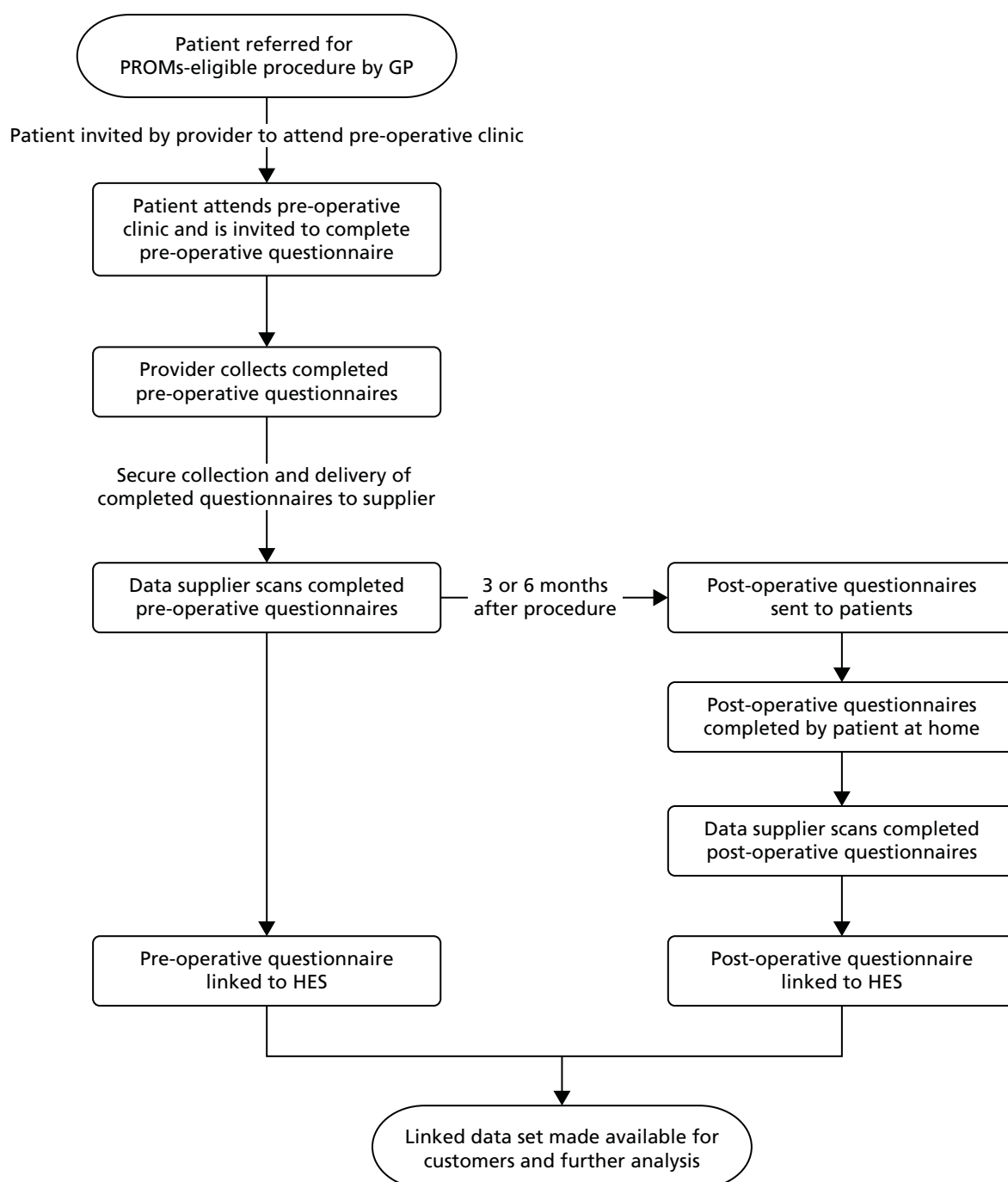


FIGURE 8 Flow chart for the collection of PROMs data in the English national PROMs programme. Reproduced from HSCIC.¹²⁰ Contains public sector information licensed under the Open Government Licence v3.0 (www.nationalarchives.gov.uk/doc/open-government-licence/version/3/).

Quality Improvement Partnership website. In addition, professional bodies have led national registries which record activity and outcomes for patients. For example, the Society for Cardiothoracic Surgery began its registry in 1977,¹¹⁰ and the National Joint Registry (NJR) was launched in 2002. However, information from these sources was not routinely made available to the public until recently. The first to do so was the Society for Cardiothoracic Surgery, which has published mortality rates for all NHS hospitals performing cardiac surgery since April 2005, following a Freedom of Information Request by the *Guardian* newspaper.¹¹¹ From 2013, the public reporting of standardised mortality rates at provider and surgeon level was also extended to include 10 further surgical specialties, including orthopaedics. As such, patients can now access data on both health gain and mortality following hip and knee replacement via the My NHS website.¹⁰⁶

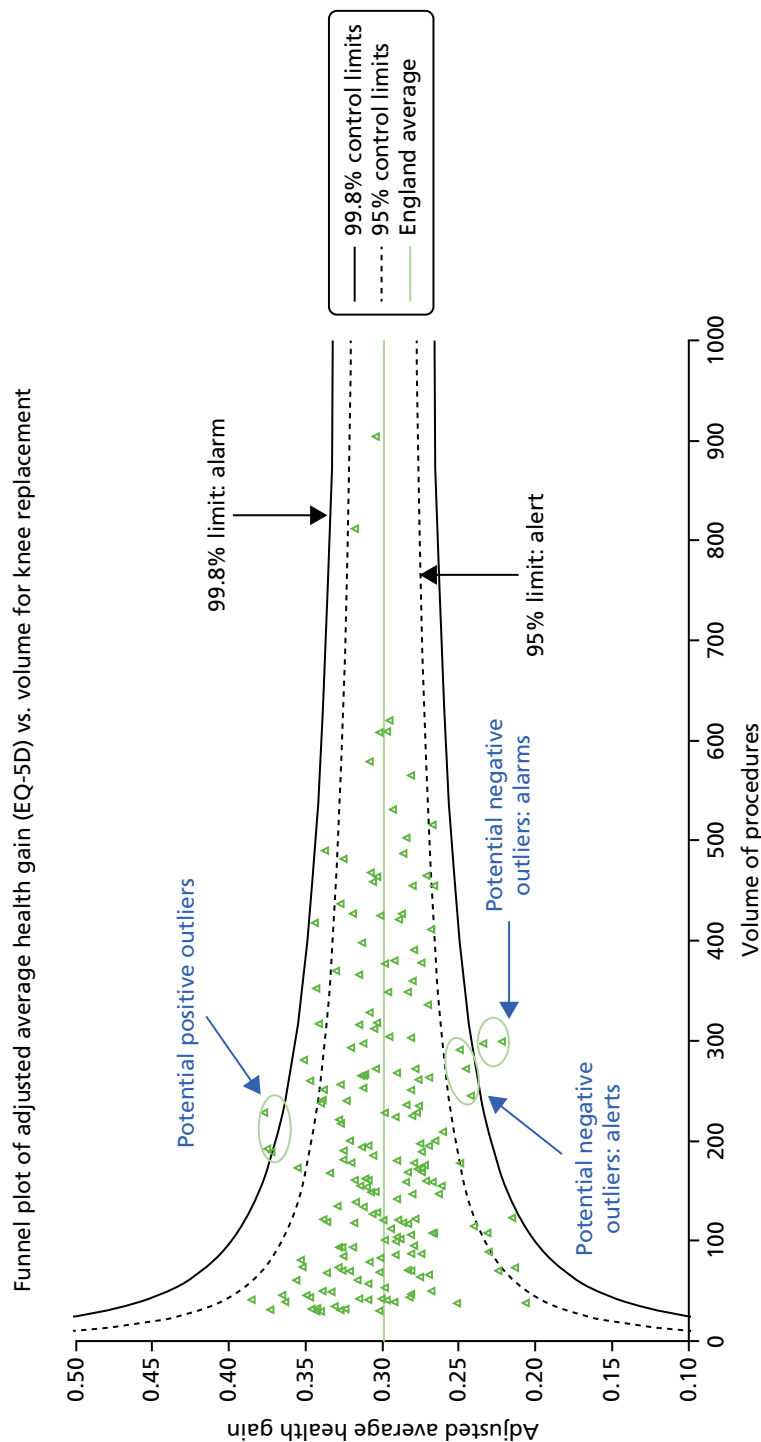


FIGURE 9 An example of a funnel plot. Reproduced from Department of Health.¹⁰⁹ Contains public sector information licensed under the Open Government Licence v3.0 (www.nationalarchives.gov.uk/doc/open-government-licence/version/3/).

In mental health services, including services provided under the government's Improving Access to Psychological Therapies (IAPT) programme, providers have been required to collect data on activity and outcomes as part of the Mental Health and Learning Disability Minimum data set since 2003.¹¹² Data are collected via an integrated mental health electronic record. Clinicians are expected to record session-by-session outcome data that are intended to both inform their clinical care of patients and monitor service performance.¹¹³ This is one example in which the same outcome data are expected to inform the care of individual patients and are then aggregated to act as indicators of service quality. Key performance indicators for IAPT services include:

- waiting times (measured as the percentage of patients waiting < 6 weeks to start therapy and the percentage of patient waiting < 18 weeks to start therapy)
- recovery rate [the percentage of patients who moved from meeting the cut-off score for 'caseness' to no longer meeting the cut-off score for 'caseness' at the end of treatment as measured by either the Patient Health Questionnaire (PHQ-9) or the appropriate anxiety disorder-specific questionnaire]
- reliable improvement (the percentage of patients who experienced a decrease in scores on either the PHQ-9 or an anxiety disorder-specific questionnaire that was greater than the measurement error of the questionnaire, irrespective of whether or not they met the criteria for caseness at the beginning or end of treatment)
- reliable recovery rate (the percentage of patients who showed either reliable improvement or recovery).

These indicators are now available to patients and the public via the My NHS website.¹⁰⁶ In addition, a summary report of a broader set of statistics derived from data collected as part of the Mental Health and Learning Disability Minimum data set are available as a report from the HSCIC website.

To support the measurement of patient experience under the NHS Outcomes Framework, the CQC launched the collection of a number of patient experience surveys across the NHS. These explore patients' experiences of a number of different services, including maternity services, inpatient services, community mental health services, accident and emergency and outpatient services, and children and young people receiving day case or inpatient care. In primary care, the GP Patient Survey examines patients' experiences of primary care. In addition, The National Cancer Patient Experience survey, which examines patient experiences of adult acute cancer services in all NHS trusts in England, has been running annually since 2010. The precise methodology for each survey differs slightly, but all involve a patient experience questionnaire being sent to a random sample of patients who received NHS services during a given time period. The findings are made available to patients via the My NHS website and the CQC website.

National surveys are run annually, and there is a time lag of several months between the collection of the data and their publication. To provide real-time feedback to hospital trusts, the Friends and Family Test (FFT) was implemented in acute hospitals in 2013. The FFT is a single question that asks patients to think about their recent experience of the service and to rate how likely they would be to recommend this service to friends and family if they needed similar care or treatment (with response options of extremely likely, likely, neither likely nor unlikely, unlikely, extremely unlikely and don't know). Data from the FFT for key services or wards (e.g. labour ward, postnatal ward, accident and emergency and inpatient wards) for each hospital are available via the My NHS website.¹⁰⁶

Financial rewards and sanctions in England

To drive QI in the NHS, a suite of financial incentives has been introduced to enable commissioners to reward providers for high-quality care. We briefly review these here to provide some contextual information about the existing reward and incentive schemes within which the national PROMs programme operates. In primary care, general practices can voluntarily participate in the QOF, which was introduced in 2004. When the QOF was first introduced, its indicators represented four main domains: clinical care, organisational, patient experience and additional services. In 2013, a new domain of public health was introduced, with some indicators from the additional services domain relocated to the public health domain; in addition, many of the indicators in the organisational domain were retired. The current QOF for 2015–16 links up to 25% of a GP's income to the achievement of indicators that reflect evidence-based care in two domains: clinical and public health.

Until 2013–14, the QOF included two indicators relating to the use of a validated depression screening questionnaire. One indicator measured the percentage of patients with depression who have an assessment of severity at the time of diagnosis, and the second indicator measured the percentage who were assessed again 4–12 weeks afterwards. However, these indicators were unpopular with GPs owing to the poor sensitivity and specificity of the instruments used (such as the PHQ-9, the Hospital Anxiety and Depression Scale and the Beck Depression Inventory),¹¹⁴ and were retired after the 2012–13 QOF owing to their ‘lack of evidence base’.¹¹⁵

For providers, the Commissioning for Quality and Innovation (CQUIN) scheme was introduced in 2009–10 to incentivise local achievement of quality targets. These targets are a set of QI goals that are agreed jointly by commissioners and providers; a national suite of targets are set each year, but commissioners and providers can also agree their own targets to reflect local issues. Each goal is accompanied by an indicator to measure achievement of the target. Currently, CQUIN payments constitute 2.5% of provider budgets. Alongside CQUIN payments, Best Practice Tariffs (BPTs) were also introduced in 2010. The purpose of BPTs is to pay prices for services that reflect the cost of best clinical practice rather than the national average cost of care. A mixed-methods evaluation of BPT, conducted in 2012, suggested that commissioners were much more aware of CQUINs than BPTs, but that providers preferred BPTs because they perceived that BPTs better reflected current evidence than CQUINs.¹¹⁶ In 2014, a BPT was introduced to support best practice in hip and knee replacement surgery. This specified that providers would be paid the tariff if they did not have an average health gain (as measured by either a generic or a relevant disease-specific PROM) significantly below the national average (99.8% significance) (i.e. they were not an alarm); had a minimum PROMs preoperation participation rate of 50%; a minimum NJR compliance rate of 75%; and a NJR unknown consent rate of < 25%. Thus, PROMs data now form an integral part of the national system for the financial reward for providers.

Summary

In this section, we have charted the history of health services performance measurement and reporting in England generally, and in the national PROMs programme specifically. PROMs are one specific form of hospital performance data, and share many of the underlying ideas and assumptions about how these interventions are intended to work. We have also mapped out some important contextual factors that not only gave rise to some of these programmes, but also shaped their implementation. These include the shift towards the use of outcomes as an indicator of the quality of health care, policy efforts to provide patient access to information on health service performance and support patient choice, and the different financial rewards and sanctions designed to motivate organisations and individuals to improve the quality of patient care. We now turn to reviewing the ideas and assumptions, or programme theories, underlying the use of PROMs and other performance data to improve patient care.

Programme theories

In this section, we focus on cataloguing the specific underlying ideas about how PROMs and other performance data are intended to improve patient care.⁸⁹ We identified a cluster of assumptions about the original intentions of the programme, followed by a series of critical voices identifying the difficulties of implementing the programme in practice, and, finally, a number of ideas about how these tensions and problems might be overcome. We have, therefore, organised our narrative along these themes. In subsequent sections, we consider how these specific theories map to abstract theories of audit and feedback, benchmarking and public disclosure.

How the intervention is intended to improve patient care

In this section, we consider the programme theories, or underlying ideas and assumptions, through which the use of PROMs and other performance data are intended to improve patient care. Berwick *et al.*¹¹⁷ theorise that the public report of performance data may improve patient care via two pathways: a change pathway, whereby providers take steps to change clinical care, and a selection pathway, whereby patients, commissioners, regulators and referring clinicians choose high-performing providers over lower-performing providers. This framework has been persuasive in shaping how the impact of performance data has been conceptualised, and it has informed the research questions and structure of a number of systematic reviews of the impact of performance data on patient care.^{19,27} Hibbard *et al.*¹¹⁸ hypothesised three different mechanisms through which the public reporting of hospital performance might stimulate QI activities. Providers are motivated to respond because:

1. Providers feel threatened by the potential loss of market share that could occur if patients decided to choose alternative, higher-performing providers.
2. Providers wish to protect their professional or institutional reputation, which may have been damaged by being publicly labelled a poor performer.
3. Providers' professional norms and values mean they are intrinsically motivated to maintain good patient care, and will take steps to improve if feedback highlights that there is a gap between their performance and expected standards of patient care.

Here, we inspect these ideas and assumptions in closer detail.

Supporting patient choice

The coalition government's White Paper *Equality and Excellence: Liberating the NHS*⁶ envisaged that PROMs and patient experience data would enable patients and their families to choose hospitals, which in turn would encourage providers to improve care, as the following quotation illustrates:

Information generated by patients themselves will be critical to this process, and will include much wider use of effective tools like Patient-Reported Outcome Measures (PROMs), patient experience data, and real-time feedback . . . feedback from patients, carers and families, and staff will help to inform other people with similar conditions to make the right choice of hospital or clinical department and will encourage providers to be more responsive.

p. 14. © Crown copyright; contains public sector information licensed under the Open Government Licence v3.0⁶

Initial guidance on the routine collection of PROMs data, issued by the Department of Health in 2008,¹¹⁹ and later documents issued in 2015,¹²⁰ also anticipated that PROMs data would be used to inform patient choice. These documents list one of the anticipated uses of PROMs data as to:

*Evaluate the relative clinical quality of providers of elective procedures . . . it can be used by patients and GPs exercising choice.*¹²⁰

Implicit in these documents is the assumption that performance data will enable patients to discriminate between poorly performing and higher-performing hospitals and that, armed with this information, they will choose higher-performing hospitals. This, in turn, will put pressure on lower-performing hospitals to improve care.

Despite PROMs data's relatively recent introduction for routine use in health care, the underlying reasoning about how they will be mobilised shares many of the assumptions of other 'feedback' or 'public reporting of performance' initiatives, such as hospital star ratings in England¹⁸ and surgical mortality report cards in

the USA. For example, writing about the assumptions underlying the public disclosure of performance data in the USA, Marshall *et al.*¹²¹ note:

In theory, making meaningful quality information available to consumers will encourage market competition based on quality and either drive out low quality providers or encourage them to improve.
Marshall *et al.*¹²¹

Others⁷⁷ have summarised this belief that market forces, created by patients choosing high-quality hospitals, will improve patient care as follows:

Market advocates believe that patients want to be able to choose between different providers; that given good information they will do so; and that these choices will be a major force driving improvements in services.

Marshall and McLoughlin⁷⁷

Furthermore, it is the public reporting of that information that is seen to produce additional pressure on providers to improve. Pinto and Pride¹²² explain that the public reporting of outcomes for cardiology procedures, such as percutaneous coronary intervention, will bring about change more quickly:

Public reporting of PCI [percutaneous coronary intervention] outcomes . . . has been implemented with the intent that patients would make educated decisions about where to get their healthcare, and providers would make practice improvements and invest in systems of care . . . hospitals and providers are more apt to rapidly adopt quality improvement measures when outcomes are publicly reported.

Reproduced with permission from Journal of the American College of Cardiology, vol. 62, Pinto, D.S. and Y.B. Pride, Paved with good intentions and marred by half-truths, pp. 416–17, © 2013, with permission from Elsevier¹²²

Thus, the theory is that patients will use PROMs and/or other performance data to compare providers and opt to receive care at higher-performing hospitals. This, in turn, will put pressure on lower-performing hospitals to improve their care, as they fear that they will lose patients and, thus, their market share.

Accountability to stakeholders

Publicising performance data was also envisaged as a means of making providers more accountable to stakeholders, including patients, who, in turn, would exert pressure on poor-performing providers to improve. For example, *Equality and Excellence: Liberating the NHS*⁶ states:

Information will improve accountability: in future, it will be far easier for the public to see where unacceptable services are being provided and to exert local pressure for them to be improved.
p. 14. Contains public sector information licensed under the Open Government Licence 3.0⁶

It is recognised that not only patients, but also regulators and government bodies, may hold providers to account. For example, Bridgewater and Keogh¹¹⁰ note that one of the key drivers for the public reporting of hospital and surgeon cardiac mortality outcomes in the UK was to 'reassure patients, their carers, hospital managers, commissioners, healthcare regulators and politicians about the quality of surgical care'.

Guidance on use of PROMs data^{119,120} also acknowledges that commissioners can use these data to hold providers to account for the quality of care they provide. In these documents, PROMs data are seen as empowering commissioners to monitor the quality of services they fund:

*Empower commissioners. PCT [primary care trust] commissioners can use the data to establish the quality of services, which they are contracting providers for.*¹²⁰

Devlin and Appleby¹²³ also argue that commissioners can use PROMs to inform the ways in which they commission services and monitor the quality of those services, including:

- monitoring the performance of the providers from whom they commission services
- specifying minimum performance on PROMs via their contracts with those providers
- incentivising providers to improve patient health by linking payment to performance on PROMs.

Similarly, the IAPT data handbook¹¹³ (p. 10) advises that outcome data on IAPT services are collected to:

- monitor the extent to which IAPT workers and services are providing evidence-based treatments that are consistently applied in the manner recommended by NICE
- assist commissioners and service providers in monitoring and improving the quality and cost-effectiveness of their services for all communities.

Here, data are used to check that services are being provided according to evidence-based guidelines, and that commissioners are receiving 'a direct return on the investment made in services' in terms of patient improvement.¹¹³ According to this theory, the process of the public, government and regulators holding providers to account for the quality of care that they provide will motivate organisations and individuals to improve patient care.

Provider benchmarking and peer competition

The Department of Health anticipated that publishing performance data would enable clinicians and managers to compare their own performance with that of their peers, and that this would instil a sense of competition between them that would provide motivation for improving performance. *Equality and Excellence: Liberating the NHS*⁶ notes:

There is compelling evidence that better information also creates a clear drive for improvement in providers. Our intention is for clinical teams to see a meaningful, risk-adjusted assessment of their performance against their peers, and this assessment should also be placed in the public domain. The Department will revise and extend Quality Accounts to reinforce local accountability for performance, encourage peer competition, and provide a clear spur for boards of provider organisations to focus on improving outcomes.

*Contains public sector information licensed under the Open Government Licence v3.0*⁶

Here, performance data are not only fed back privately to providers but also placed in the public domain. This comes with the implicit expectation that the public will respond by exerting pressure on poor-performing hospitals to improve, as described previously (see *Accountability to stakeholders*) and that improvement will also occur through providers competing with each other.

The most recent version of the PROMs guidance¹²⁰ also envisages PROMs data enabling clinicians, managers and commissioners to benchmark their performance:

Assessing the relative clinical quality of providers of elective procedures; for clinicians, managers and commissioners benchmarking their own performance.

*p. 8. Contains public sector information licensed under the Open Government Licence v3.0*¹²⁰

This places responsibility on providers and managers to use PROMs data to monitor *their own* performance and, implicitly, to then take steps to improve this. Devlin and Appleby¹²³ argue that PROMs data should:

... act as a focus and a starting point for providers, first to identify the reasons for their performance, and then to identify what they need to do in order to improve.

*Devlin and Appleby*¹²³

The mechanism through which this is hypothesised to occur is that providers are motivated to improve their practice in order to be as good as or better than their peers. The private sector provider Bupa (now owned by Spire) pioneered the routine collection of PROMs data to monitor the outcomes of surgery some years before this initiative was introduced into the NHS.¹²⁴ Its then chief executive, Andrew Vallence-Owen, was interviewed for a *British Medical Journal (BMJ)* article¹²⁵ and argued that QI arose because surgeons competed with each other, not for patients, but for the professional prestige associated with having better outcomes than their peers:

Doctors are quite competitive . . . Once they see each other's data they want to do as well or better. So you get continuous quality improvement out of this.

Reproduced from NHS goes to the PROMS, Timmins N, vol. 336, p. 1465, © 2008 with permission from BMJ Publishing Group Ltd¹²⁵

In other words, it is the provision of comparative, benchmarked data to enable surgeons to compare themselves with their peers which is important in stimulating QI.

There is some debate about whether private feedback of benchmarked performance would be sufficient to motivate change or whether there is an added benefit of publicising this performance. Marshall *et al.*¹²¹ note that the public reporting of performance data 'is based on the assumption that organisations and professionals have an intrinsic desire to improve practice' but may be prevented from doing so owing to lack of time, knowledge and resources and competing priorities. Thus, reporting the data publicly is intended to 'increase providers sensitivity to their performance by reminding, refocusing or shaming them into action'.¹²¹ The theory is that the private feedback of performance data to providers, for example via local or national audits, has not been sufficient to motivate providers to act on these data. Putting the data into the public domain is seen as a means of shaming providers into action by publicly labelling them as poor performers. According to this theory, it is reputational damage that motivates providers to take steps to improve patient care.

The role of the media

Several commentators have noted that the media were instrumental in the move towards publicly disclosing provider performance data. For example, The Health Care Financing Administration (HCFA) reports were one of the first reporting systems in the USA. The reports were first released in 1986, and were originally designed for use only by 'state peer review organisations', that is, organisations contracted by the federal government to review the quality of medical care delivered to Medicare-funded patients.¹²⁶ They were not intended for use by the public. However, as their existence became known to consumer groups and the press, these groups called for their release under a Freedom of Information Act and the reports were then made public in December 1987.¹²⁶ The reports themselves were subsequently criticised for poor case-mix adjustment and shelved (an issue we will return to in subsequent sections). In the UK, the *Guardian* newspaper published named surgeon mortality data following a Freedom of Information request. Bridgewater and Keogh¹¹⁰ argue that, in the UK, such media-led initiatives prompted a professionally oriented model of public disclosure, in which professional organisations such as the Society for Cardiothoracic Surgery worked with the Healthcare Commission to publish surgical results. Bridgewater¹²⁷ observed that:

[W]ithout the involvement of the press, it seems unlikely that named surgeon outcomes for cardiac surgery would be in the public domain.

The role of the media in publicising performance data has also been the subject of much debate. One theory is that the media can be used to ensure that the report reaches the intended stakeholders. For example, guidelines on how to maximise the impact of public reporting issued by the Agency for Healthcare Research and Quality¹²⁸ advise those involved in public reporting initiatives to harness the power of the media in getting their report publicised. They advise issuing press releases and holding new conferences, but recognise that these alone may not be sufficient to obtain the depth, breadth or quality of coverage required to raise awareness of public reports.

However, it has also been recognised that the media's interests lie in selling newspapers, and thus they may be more inclined to report 'bad news' stories, or to misrepresent or overly simplify complex information, such that it does not represent a true picture of performance. The Agency for Healthcare Research and Quality guidelines¹²⁸ also recognised this, and warn that a shocking story about a patient's experience of poor care in a specific hospital is more likely to be reported in the media than a news item about the release of a report on hospital quality.

In turn, others have highlighted that misrepresenting or misunderstanding the information contained in a hospital report can have a negative impact on hospitals and staff:¹²⁹

There is no doubt that public disclosure may have negative effects for hospitals and their staff. If the reasons to criticize the hospital . . . are due to misrepresentation or lack of ability to understand, then the message the disclosure brings is of no use or is detrimental.

Mebius¹²⁹

To counteract the possibility that the media may distort or misrepresent the data, the Agency for Healthcare Research and Quality guidelines¹²⁸ suggest actively working with the media to ensure that the correct message is communicated, and ensuring that stakeholders provide a consistent message. They instruct that one way of achieving this is to create guidelines for interactions with the media to ensure that all stakeholders provide a consistent message about the public report.

Assessing the appropriateness of surgery

Initial guidance on the routine collection of PROMs data was issued by the Department of Health in 2008.¹¹⁹ This document outlined a number of potential uses of PROMs data, including using the data to evaluate the appropriateness of referral for surgery:

Assess the appropriateness of referrals to secondary care. PROMs data can be used to establish whether referrals for elective procedures are appropriate by examining variation in baseline PROMs scores across the country.

Contains public sector information licensed under the Open Government Licence v3.0¹¹⁹

However, later guidance¹²⁰ on the use of PROMs data to assess the appropriateness of referrals has also been replaced with:

Assessing the relative health status before operations: PROMs data will provide a measure of patients' self-reported health status before they undergo operations. Exploring variations in this data along with other indicators of surgery, such as the patients' social situations, other medical conditions and risk of deterioration and/or complications, could establish benchmarks.

Contains public sector information licensed under the Open Government Licence v3.0¹²⁰

Here, there is a shift away from any reference to assessing the appropriateness of surgery towards simply observing variations and establishing benchmarks. This change in emphasis could reflect reaction to the responses to an analysis of the pilot PROMs data carried out by Devlin *et al.*,¹³⁰ which found little health gain, on average, measured by the EQ-5D following surgery or varicose vein repair. This was reported in the *Health Service Journal*⁷⁶ as suggesting 'at least £144 m is being spent annually on carrying out operations on people who either have no significant complaints about their health before surgery or report that their condition is changed or worse afterwards'.

Black⁶⁶ highlights that the misuse of PROMs data to restrict access to elective surgery is one of the possible unintended (but not unforeseen) consequences of the PROMs programme and advises that it is not possible to identify patients who will not benefit from the procedure preoperatively.

Summary of how patient-reported outcome measures are intended to improve patient care

Here we have summarised the key theories underlying how and why PROMs and other performance data are intended to improve patient care. This analysis suggests that there are numerous competing but not necessarily mutually exclusive mechanisms through which organisations and individuals might be motivated to improve patient care: through fear of losing market share, a desire to be as good or better than peers, concerns about damage to professional reputation and pressure from stakeholders to be answerable for the quality of care provided. However, we also uncovered a number of concerns that publication of performance data may not achieve these desired outcomes for a number of different reasons. It is these concerns that we turn to next.

Why the intervention may not work as intended

There were four distinct 'cautionary tales' among the programme theories in our analysis. The first set of theories questioned the proposed mechanisms through which PROMs were expected to work, in particular the idea that patients were aware of, able to make sense of and would act on performance data to choose hospitals, and that market competition would improve patient care. The second set related to the potential unintended or adverse consequences of PROMs data. The third set related to the nature of the data and their analysis, with concerns about threats to data credibility, and a fourth set clustered around characteristics of the data that may make them more or less 'actionable' by providers. We consider each set of ideas in turn in this section.

Does patient choice drive improvements in patient care?

Numerous commentators have questioned the assumptions underlying the idea that market forces, driven by patients choosing higher-quality hospitals, will improve patient care. The literature in this area is voluminous and here we are able to highlight only the key points from across the terrain. A cornerstone of neo-classical market economics is that individuals operating within it are rational and seek to maximise the value or 'utility' that they derive from goods and services (health and otherwise). This theory makes strong assumptions about agents in the market, for example that they have consistent and robust preferences for services, and that they have the relevant information to make informed choices. In countries where market forces are present in the health-care sector (e.g. the USA), patients may make choices between providers based on information about both quality and price, and we might expect to see some relationship between quality indicators and market share. In countries where choice is available but competition is based only on quality (e.g. England), the relationship will be attenuated. That said, the costs (e.g. additional travel costs) of selecting a different provider may act as a disincentive. In attempting to maximise utility in choosing a provider, individuals will take into account a number of factors, only one of which will be quality.¹³¹

The market share theory is further attenuated by the assumptions of profit-maximisation and competition. The former assumes that health-care providers operate with the aim of maximising profits or income but – particularly in health care – this may not be the case. When individual health-care professionals make decisions regarding health-care provision, they are likely to aim to maximise the utility for the patient (or group of patients) rather than profits. Furthermore, even when profit or income is a motive, increased market share may not deliver this if the marginal cost of an additional patient exceeds the marginal revenue. Thus, even when patients are able to make an informed choice of provider, the provider may not accept increased patient through-flow.

Many of the theories and assumptions underpinning the hypothesised market mechanism have been found not to reflect how individuals and firms operate in reality. Behavioural economists have offered a number of competing theories for economic and choice behaviour. For example, although an individual receives negative information about a current provider, he or she may be unwilling to switch provider due to the status quo bias.¹³² This unwillingness might arise owing to uncertainty about whether a new provider will be better, and concerns about loss aversion.

Further criticisms question the fundamental assumption that patients behave like rational maximisers, and argue that their capacity to engage with the choice agenda is heavily shaped by context.¹³³ For example, Gabe *et al.*¹³³ argue that many patients are not aware that they have any choice about where they can be treated or that there may be variations in the quality of care between different hospitals. Others maintain that patients do not seek out quality information, do not understand it and do not use it. For example, Marshall and McLoughlin⁷⁷ observe:

Although patients are clear that they want information to be made publicly available, they rarely search for it, often do not understand or trust it and are unlikely to use it in a rational way to choose the best provider.

Marshall and McLoughlin⁷⁷

Instead, they argue, the underlying mechanism driving improvements following feedback of performance data are provider attempts to maintain their reputation. Furthermore, Kullgren *et al.*¹³⁴ argue that patients do not use information from report cards but instead rely on information from their personal social networks to make their choice:

*[M]any patients currently choose their providers primarily, if not exclusively, based on the endorsement of trusted personal sources, such as their social networks or physician referrals.*¹³⁴

Kullgren *et al.*¹³⁴

Others have highlighted that patients vary widely in both their desire and their capacity to make choices. For example, Fotaki¹³⁵ notes:

... although some patients may wish to choose an alternative provider and some can benefit from participating in choices about their treatment, many may not wish to make them, or may be unequally equipped to exercise such choices.

Fotaki¹³⁵

Doubts that patient choice and market forces will drive improvements in patient care now appear to have been taken on board by the current Secretary of State for Health, Jeremy Hunt. The *Health Service Journal*¹³⁶ reported an interview with Jeremy Hunt in November 2014, following the publication of the *Five Year Forward View*. In this interview, Jeremy Hunt is quoted as indicating that patient choice is unlikely to be the main driver of performance improvement because patients remain very loyal to their local hospitals, even when such hospitals have well-reported care failings. Instead, Jeremy Hunt highlighted that openness about quality using CQC ratings and other data would lead to improvement, as organisations and professionals comparing themselves with each other would, in turn, experience a strong desire to improve. Thus, the assumption here is that patients do not always exercise choice when performance data indicate poor-quality care, but instead remain loyal to local hospitals. Performance data are perceived to drive improvements via peer comparison and an intrinsic desire to improve care, rather than through patient choice and market forces.

Unintended or adverse consequences of performance data

Several commentators have highlighted the potential unintended consequences of publishing performance data that may militate against their intended effects of improving the quality of patient care.^{137,138} For example, Braithwaite and Mannion¹³⁹ caution that, even when the technical problems of performance management systems have been addressed, these systems can have little meaningful impact on performance because these data are:

[I]gnored, argued over, politicised, improvement efforts flounder, or targets and indicators have perverse effects. Indeed, public performance measures are not neutral assessments of performance, but can alter behaviour in unintended and dysfunctional ways.

Braithwaite and Mannion¹³⁹

Specific examples of unintended consequences include the manipulation or ‘gaming’¹⁴⁰ of data through a range of activities that give the appearance of improved performance but involve no real change in the underlying performance. This could include purposefully falsifying data but often involves more subtle behaviours such as changing the way data are coded or recorded and modifying the definition or interpretation of the indicators.

Another criticism is that the publication of performance data leads to distorted priorities or ‘tunnel vision’,¹⁸ where organisations focus on the areas of care measured by the performance data to the detriment of other important, but unmeasured, areas of care. Marshall *et al.*¹²¹ note:

The unintended consequences of publication have been highlighted, including manipulation of data and an inappropriate focus on what is being measured, to the detriment of other areas of activity.

*Marshall et al.*¹²¹

Economic theorists Holmstrom and Milgrom¹⁴¹ describe this type of behaviour as ‘effort substitution’, which arises when an indicator focuses on the aspects of an agent’s work that are easiest to measure (e.g. the quantity of outputs) rather than aspects that are harder to measure (such as quality of output). The indicator itself serves to focus the agent’s attention on aspects of their work that are measured at the expense of those aspects that are not. Bevan and Hood¹⁴² also argue that performance indicators can have ‘threshold effects’, whereby pressure is put on poor performers to improve, but those doing better than the target have the perverse incentive to let their performance deteriorate towards the target. We discuss this in more detail in subsequent sections.

Bevan and Hood¹⁴² consider threshold effects, effort substitution and gaming collectively as ‘gaming’, defining it as ‘reactive subversion such as “hitting the target and missing the point” or reducing performance where targets do not apply’. They hypothesised that the extent of gaming under any performance indicator system depends on both the motivation and the opportunity stakeholders may have to massage the data.

These theories, largely drawn from economics, characterise providers as actively engaged in a process of deliberate subversion of the performance indicator system. In contrast, sociologists have drawn attention to the social organisation of the production of such indicators.¹⁴³ From this perspective, the distortion of indicators occurs not as a deliberate attempt to manipulate the system, but as a result of clinicians attempting to make sense of inherently subjective and messy criteria in the face of uncertainty.¹⁴⁴

Another potential unintended consequence is for clinicians to avoid treating sicker patients to avoid poor publicly reported outcomes. For example, in a news bulletin following the announcement of plans to extend the publication of surgical mortality rates to other specialties, the chairperson of the British Medical Association’s Consultants Committee expressed concern that:¹⁴⁵

Some surgeons are deterred from taking on very complex, high-risk procedures because published simplistic league tables count against them.

Whether or not this unintended consequence is substantiated by objective data is contested; here we merely note the conjecture that it may occur.

Summary of unintended consequences

Performance data may give rise to a number of unintended consequences. These include gaming data to give the appearance of changes in performance, without real changes in performance; effort substitution, where efforts are directed to improving what is measured at the expense of what is not; threshold effects, where poor performers feel pressure to improve, but high performers may allow performance to deteriorate towards the mean; and avoidance of treating sicker patients in order to improve performance data. Whether or not and how these effects occur are contested; it will be for our evidence review to ascertain when and why they occur.

Data credibility and attribution: do performance data reflect the quality of care?

A collection of theories about the use of performance data wrestle with the question of whether or not the data actually reflect the quality of patient care and the validity of the 'outlier' label. The basic theory here is that if clinicians do not perceive the data as credible, they will not use them to initiate changes in the quality of clinical care. The sources of the difficulty with PROMs data reflecting the quality of care, or their lack of credibility, are seen to arise owing to the accuracy of the data on which the indicator is based, disputes about the adequacy of case-mix adjustment, the timing of outcome measurement and the level at which performance data are reported. These are now discussed in turn.

Data accuracy

This theory argues that the success of public reporting depends, in the first instance, on the accuracy of the underlying data used to produce performance reports or indicators. In the USA, a particular bone of contention is the formulation of indicators based on routinely collected administrative data gathered by insurance companies at patient discharge to bill payers, which are deemed by many clinicians to be inaccurate, versus the use of data extracted from patient notes by hospital representatives, requiring additional resources to obtain. Similarly, in the UK, many have questioned the accuracy of Hospital Episode Statistics data, either to produce indicators themselves, or in the case of PROMs data, to link PROMs questionnaires to patient episodes. For example, Naylor¹⁴⁶ comments:

The struggle with valid description of processes and outcomes of clinical care will continue as long as health services researchers and private profilers are forced to dredge administrative databases. Such databases have well recognized limitations in characterizing patients, clinicians and institutions.

Naylor¹⁴⁶

The perceived difficulty is that these databases were originally developed for a different purpose, that of enabling hospitals to bill insurers or commissioners for care provided, rather than for QI purposes.

Case-mix adjustment

This theory addresses the question of whether or not case mix adjustment can ensure that performance data provide a valid indicator of the quality of care provided by an organisation. It is assumed that performance data reflect the quality of care provided by a hospital and will enable differentiation between 'poor-performing' and 'well-performing' hospitals. However, patient outcomes are determined not only by the quality of care provided by the hospital, but also by the patient's baseline health, comorbid conditions and other patient characteristics, which are beyond the control of the hospital. Therefore, to represent the outcome that can be attributed to the care provided by the hospital and, thus, enable meaningful comparisons between hospitals, outcome data are adjusted for variations in patient characteristics that may influence the patient's likely benefit from the care they receive, but are not under the control of the hospital. This is known as case-mix or risk adjustment. Black⁶⁶ highlights the need for 'sufficiently robust adjustment for differences in case-mix to achieve credibility' for PROMs data. Similarly, Marshall and Brook¹⁴⁷ argue that the risk of providers refusing to treat disadvantaged groups can be attenuated by 'careful adjustment of risk and case-mix'.

The precise model that should be included in risk adjustment of performance data has, however, been the subject of much debate. Even using the same variables and data, different models can produce different findings. Iezzoni *et al.*¹⁴⁸ compared 10 different, commonly used risk adjustment models for mortality data on 100 hospitals. They found only fair to good agreement between different methods of case-mix adjustment and concluded:

For an individual hospital, perceptions of mortality performance could vary according to different severity adjustment methods . . . [which raises] . . . important questions for report card efforts to judge hospital performance by use of severity-adjusted death rates . . . it is important to weigh what actions may be reasonably founded on this information.

Iezzoni *et al.*¹⁴⁸

Similarly, Grant *et al.*¹⁴⁹ found that the choice of risk assessment model in the analysis of cardiac surgeon specific mortality data influenced which surgeons would be identified as outliers at 95% control limit.

Others have also warned against the use of outcomes data as an indicator of hospital performance because case-mix adjustment is always imperfect and can never guarantee an unbiased comparison between providers, as there will always be unmeasured prognostic factors that may influence outcomes. For example, Lilford *et al.*¹⁵⁰ argue:

Therefore, even if an agreed risk-adjustment method could be derived, outcomes could still vary systematically between providers because we can never be sure that risk adjustment is not hampered by unmeasured prognostic factors . . . Making judgements about the quality of care on the basis of risk adjusted comparisons cannot guarantee that like is being compared with like.

Reprinted from The Lancet, Vol. 363, Lilford R, Mohammed MA, Spiegelhalter D, Thomson R, Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma, pp. 1147–54, © 2004, with permission from Elsevier¹⁵⁰

Others acknowledge that the 'perfect' method of case adjustment is unachievable but argue that this should not prevent the publication or feedback of performance data. Instead, they advocate that the credibility of data can be enhanced if those responsible for publishing the data are engaging in an ongoing process of improving the data's quality. As Marshall and Brooke¹⁴⁷ argue:

It is important for those who publish data to show a commitment to investing in the process and progressively improving the quality of the data and the validity of comparisons arising from the data. However, it makes little sense to 'wait for better data' – data will always be imperfect and, as one commentator stated, it is important not to let 'perfect be the enemy of the good'.

Marshall MN and Brook RH. Public reporting of comparative information about quality of healthcare. Med J Aust 2002;176(5):205–206. © Copyright 2002 The Medical Journal of Australia – reproduced with permission¹⁴⁷

Still others have argued that the risk of misrepresentation as a result of imperfect case-mix adjustment can also be reduced through the use of multiple different forms of performance data. For example, Naylor¹⁴⁶ contends that:

Despite the best risk-adjustment algorithms, clinical outcomes are often confounded by factors other than [the] technical quality of care or clinical decision making. Well balanced performance profiles should include a sensitive array of process-of-care measures, carefully chosen clinical outcomes, and patient perception and satisfaction surveys.

Naylor¹⁴⁶

He also reasons that involving all stakeholders in the development of performance indicators can stave off concerns about their credibility. He argues that the development of performance indicators need to be:

[I]nformed by meticulous analytical methods that do everything . . . to level the playing field, eliminating unfair comparisons . . . by working closely with clinicians, hospitals, and other healthcare providers from the outset, quality analysts can forestall criticism of the credibility of their reports and catalyse positive change rather than professional defensiveness.

Naylor¹⁴⁶

Although not explicitly stated by Naylor, we can hypothesise that involving stakeholders may increase the credibility of the data through ensuring that the case-mix algorithms are appropriate, through increased ownership of the data, or through both.

When to measure

This theory argues that the choice of time point at which to measure outcomes following a procedure may either fail to capture the full benefit of the procedure on patient outcomes or become confounded by factors not related to the procedure. Black⁶⁶ acknowledges that 'judging the best time to assess outcome after an intervention so as to be able to attribute it to that intervention is also contentious'. For knee and hip replacement, postoperative PROMs data are collected 6 months after the operation, and for groin hernia repair and varicose vein surgery, they are collected 3 months after the operation. As Browne *et al.*¹⁵¹ observe:

The primary purpose of the English PROMs programme is the detection of deviant performance by surgical providers. Implicit in the methods is the assumption that the performance of these providers can be fairly judged at six months after surgery as the clinically important benefits of surgery have accrued. If this is not the case it is possible that some providers are being unfairly assessed.

Reproduced from Browne et al.¹⁵¹ under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>)

Browne *et al.*¹⁵¹ note that the cut-off point of 6 months to measure the impact of hip and knee replacement was based on 'clinical consensus' and on the need for the output to be timely to enhance the likelihood that it will stimulate health-care providers to review and improve the quality of their care. By choosing 6 months, it is not assumed that a patient has reached the maximum benefit of the operation; however, some have expressed concern that the use of 6 months as the cut-off point does not capture the longer-term benefit of the interventions.¹⁵²

Maynard and Bloor¹⁵² argue that capturing outcomes too long after a procedure makes it difficult to disentangle the relative contribution of the procedure from other aspects of care that may contribute to the patient's functioning:

Recording 'success' after three or even six months [after a procedure] may be incomplete. To go beyond this time period risks factors other than the procedure being evaluated affecting outcome.

Maynard and Bloor¹⁵²

Their example points to the issue regarding what aspects of care PROMs data are intended to reflect: the skill of the individual surgeon in performing the surgery, the teamwork and co-ordination of the operating team working with this surgeon, the quality of care provided immediately postoperatively (e.g. pain relief) or the quality of community care received by patients in the community in the months following surgery (e.g. from physiotherapists) – or all of these.

Attrition bias

When using PROMs as the outcome measure, it is necessary to have a baseline assessment (pre treatment), as this is the most powerful predictor of the post-treatment PROM, as well as a post-treatment assessment. This adds to the complexity, burden and cost of employing PROMs to assess the quality of care. In addition, the need for two data collection points means that there will be some attrition through patients not responding to the post-treatment PROMs, although, following major interventions such as hip replacements, the response rates are very high. Work by Hutchings *et al.*⁶⁷ revealed that older, sicker patients were less likely to complete PROMs, and they argue that if non-response is associated with outcome, then rates of non-response to postoperative questionnaires would need to be taken into account when these measures are being used to compare the performance of providers or to evaluate surgical procedures. Furthermore, current data on the HSCIC website indicate considerable variation in participation rates between providers, suggesting wide differences in whether or not and how providers administer questionnaires and manage the process of completion. Evidence suggests that providers with lower participation rates are more likely to be erroneously classified as an outlier.¹⁵³

Level of analysis

There is also considerable debate about the level of analysis at which PROMs and other performance data should be published. As Black⁶⁶ notes, currently PROMs data are provided at the level of the hospital, with

the implication that it is variation in the care between hospitals that gives rise to variation in outcome. Some have argued that this does not enable patients to distinguish between the quality of individual surgeons, and that surgeon-level data are required. However, others have questioned the utility of surgeon-level data, as the numbers may be too small to permit meaningful comparisons between surgeons, and because surgeons work in teams and therefore outcomes reflect the work of the team and not just of the surgeon. As Sir Bruce Keogh, medical director of the NHS, argued in an interview in the *BMJ* in 2008:¹²⁵

Professor Keogh says he does not believe individual level consultant data should be published yet and probably not 'for some time down the line' not until people are convinced that the data are accurate, and even then only if the individual data are meaningful. Doctors increasingly work in teams, he notes, so individual data may not always tell the story.

Reproduced from NHS goes to the PROMS, Timmins N, vol. 336, p. 1465, © 2008 with permission from BMJ Publishing Group Ltd¹²⁵

Trusted source

This theory highlights that credibility in the eyes of recipients is influenced not just by the data themselves, but also by the perceived 'trustworthiness' of the body supporting or initiating data collection and the perceived drivers of this process. For example, Marshall and McLoughlin⁷⁷ highlight a number of conditions under which the use of performance data can be optimised, including the idea that the data are perceived as coming from a 'trusted source':

It is important that users perceive the information as coming from a trusted source. Information providers that might be regarded as having ulterior motives, such as government or for-profit organisations, may not be perceived to be independent. Partnerships between the health service, and respected professional bodies and academic institutions may be seen as more trustworthy.

Reproduced from How do patients use information on health providers?, Marshall M, McLoughlin V, vol. 341, p. 5272, © 2010 with permission from BMJ Publishing Group Ltd⁷⁷

It is argued here that the perceived motives of the body promoting the collection and use of performance data can influence the degree to which the data are trusted by recipients. Wolpert¹⁶ also shares this view and highlights the tensions that have emerged between the collection and feedback of PROMs data to support clinical decision-making for individual patients and the use of PROMs data for performance management in the context of mental health services. She argues that the perceived primary driver for the introduction of PROMs data collection in clinical practice influences the degree of clinical engagement with the process:

It may be crucial that in introducing PROMs into clinical practice frontline clinicians are introduced to the tools through a prism of collaborative working and shared decision making rather than as tools primarily used for audit or performance review . . . an underlying ethos of collaborative working and shared decision making, and a focus on using PROMs as part of clinical conversations, promotes greater clinician engagement and willingness to trial the use of PROMs.

Reproduced under the terms of the Creative Commons Attribution License, which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited¹⁶

Similarly, Marshall and Brooke¹⁴⁷ argue that clinical engagement in selecting the indicators may also increase engagement with the collection and dissemination of performance data:

. . . we know that forcing new initiatives on reluctant professionals is not the most effective way of changing attitudes, and the introduction of report cards is more likely to be successful if doctors are encouraged to take a lead, particularly in selecting the performance measures.

Marshall and Brooke¹⁴⁷

The World Health Organization¹⁵⁴ also points out that hospitals often require incentives to participate in public reporting programmes. It observes that the perceived rationale and possible benefits and risks of collecting performance data can influence the extent to which other incentives are required to promote participation in such schemes. It argues that if public reporting programmes are perceived by staff to have intrinsic value to them (e.g. through enabling team building, supporting clinical or professional development or assisting with risk management), then hospitals have less need for financial or market incentives to participate. However, they also note that staff are unlikely to be motivated to participate if performance data may lead to public blame or litigation.

Thus, these theories suggest that it is not just the validity of the data itself that can influence how providers respond, but also the perceived rationale for the data's collection and whether or not providers feel that they provide information relevant to their own work.

Summary of data credibility

The theories reviewed here revolve around that idea that, unless providers perceive PROMs or other performance data to be credible and accurately reflecting the quality of patient care, they will not trust the data and, consequently, will not take steps to improve patient care. These arguments suggest that providers' trust in the data rests not just on the technical aspects of validity – that is, case-mix adjustment, the level of analysis and timing of assessment (although these are important) – but also on the perceived motivations behind data collection. These theories also highlight potential tensions between the use of the data for performance management and the use of the data to inform the clinical care of patients. They suggest that clinicians are more likely to engage with performance data if data are provided that can also be used to inform the clinical care of patients.

Actionability

A fourth set of theories addresses the ideas and assumptions about whether or not and how different stakeholders are enabled to interpret performance data and identify the cause of poor care and, subsequently, develop solutions. These focus on the timeliness of the data, the ways in which the data are presented and the skills and capacity of the organisation to interpret the data, whether or not different forms of performance data are better able to guide the selection of QI strategies and, finally, whether the purpose of feedback is to improve the performance of poor performing organisations or all organisations.

Timeliness

One theory as to why performance data may not lead to improvements in patient care is because of the time it takes to collect data, adjust for case mix and then disseminate. If prolonged, the data may reflect care provided in the past, rather than that provided currently. This, in turn, makes it difficult to identify and then address current areas of poor care. For example, in a paper that attempted to understand why 'public reports have not had more impact on system performance', Guerriere¹⁵⁵ speculated that:

The impact of healthcare performance reporting may be diminished by the age and quality of the reports. Getting old news about problems you had a year ago does little to motivate change today.

Guerriere¹⁵⁵

Similarly, in a paper debating the potential outcomes of the Australian government's plans to introduce performance data into its health-care system, Braithwaite and Mannion¹³⁹ warned that:

Performance measurement systems are necessarily back-ward looking, as it takes time to assemble and disseminate data. By the time a problem is spotted, it may be too late to do anything about it.

Braithwaite and Mannion¹³⁹

A suggested solution is the feedback of 'real-time' performance data: that is, feedback in which the delay between data collection and feedback is minimised. Such an approach has been introduced in England, with the real-time collection and feedback of one form of patient experience data, the FFT, implemented

in acute hospitals in 2013. The FFT is a single question that asks patients to think about their recent experience of the service and rate how likely they would be to recommend that service to friends and family if they needed similar care or treatment (with response options of extremely likely, likely, neither likely nor unlikely, unlikely, extremely unlikely and don't know). Recent NHS England guidance¹⁵⁶ on the FFT highlights that its timeliness means that it has some advantages over traditional survey methods of collecting feedback:

Compared to traditional survey methods, where there is often a considerable time-lag between the collection of feedback and the survey results, the FFT is a timely feedback tool. This can help providers to understand their areas of strength and weakness – and drive improvements in patient care – very quickly.

p. 18. Contains public sector information licensed under the Open Government Licence v3.0¹⁵⁶

However, NHS England acknowledged that real-time feedback also brings disadvantages, such as the loss of the ability to benchmark performance with other providers:

The FFT does not provide results that can be used to directly compare providers because of the flexibility of the data collection methods and the variation in local populations. This means it is not possible to compare like with like. There are other robust mechanisms for that, such as national patient surveys and outcome measures. The FFT can help mark progress over time for organisations and still provides patients with useful data to inform choice.

p. 18. Contains public sector information licensed under the Open Government Licence v3.0¹⁵⁶

Here we see the tensions that arise between the timeliness of the data and their credibility and usefulness. The simplicity of the indicator (a single question) and the quick collection and feedback of these data, which ensure that these data are timely, also mean that there is not sufficient time to adjust these data for case mix, and, thus, ensure that like is compared with like, or to present these data in a format that allows comparison with other providers.

Presentation and interpretation of the data

It has also been speculated that the ways in which performance data are presented influences whether or not their intended recipients are able to interpret and make sense of the data and, in turn, take action in response to these data. The most appropriate way of presenting PROMs and other performance data to facilitate accurate interpretation has been the subject of much debate. For example, in 2009, Black and Jenkinson¹⁵⁷ argued against the use of rank ordering or 'league tables' of providers and in favour of the use of funnel plots:

Ranking of providers by score is also problematic because the performance of most providers does not differ much. Such rank orders may have little meaning and can give an artificial distinction between those at the top and those at the bottom. It is better to present data in funnel plots . . . as it avoids ranking and instead focuses attention on any outliers that require local in-depth investigation to determine the reasons for their poor performance.

Reproduced from Measuring patients' experiences and outcomes, Black N, Jenkinson C, vol. 339, pp. 202–5, © 2009 with permission from BMJ Publishing Group Ltd¹⁵⁷

The ways in which PROMs data have been fed back to providers, managers and commissioners has been criticised for being difficult to interpret. For example, a *BMJ* review¹⁵⁸ of the PROMs conference held in 2013 cites Professor Nick Black expressing a wish that the group assembled to advise the Department of Health on the PROMs programme would discuss how the presentation of PROMs data could be improved:

Black said that after a hiatus caused by the government's reorganisation of the NHS in England, the PROMs programme was now moving ahead, and the first meeting of a PROMs Advisory Committee was planned by NHS England to be held soon. Among other things he hoped that the committee,

of which he will be a member, would discuss the presentation of PROMs data by the NHS Information Centre. 'All attention has been to input, none to output' he said. 'The spreadsheets published by the Information Centre are a complete turn-off. At least 20% of the budget should be spent on outputs.'

Reproduced from Patients' rating of treatment tells you more about patients than hospitals, research concludes, Hawkes, N, vol. 347, pp. f6916, © 2013 with permission from BMJ Publishing Group Ltd¹⁵⁸

Furthermore, performance data have a wide range of intended audiences: clinicians, commissioners, GPs, NHS managers and also patients. It is unlikely that the same method of presenting performance data will meet the needs of each of these different audiences. One proposed solution is to make use of web technologies that allow different audiences to select information they want and how it is presented; as Marshall and McLoughlin⁷⁷ note:

New web technologies, in particular could be used to enable users to define what information they want and how they want it presented. This would help to meet the needs of disparate audiences and recognises the heterogeneity among potential users.

Marshall and McLoughlin⁷⁷

Another purported solution to this problem is to simplify the presentation of complex data by combining data into a single score or by using simple 'star ratings'. However, this approach creates another set of difficulties. Naylor¹⁴⁶ argues that simple ratings do not convey uncertainty in the indicator (e.g. as expressed by confidence intervals) and may lead to exactly the problem (noted above by Black and Jenkinson) of creating an artificial or inaccurate distinction between hospitals:

Communication of complex performance data is difficult. Healthgrades.com uses a 'star system' to improve public accessibility, and many other agencies use similar qualitative markers to identify higher and lower performing providers. But simplified rating systems preclude the acknowledgement of uncertainty in rankings and may misclassify 'borderline' hospitals.

Naylor¹⁴⁶

However, others also recognise that the appropriate presentation of data is only one element in the process of using data to bring about change. Guerriere¹⁵⁵ argues that organisations require skills in interpreting performance data, but also the ability to translate that information into operational improvements:

A key issue is the lack of understanding throughout the system of what to do with performance information . . . Without the skills needed to interpret available performance data in order to effect operational improvements, performance reports are destined to have a very limited impact.

Guerriere¹⁵⁵

This leads us to our next set of theories, which consider whether or not performance information has sufficient 'diagnostic' information to direct organisational change.

Performance data as judgement or as a catalyst for investigation

This set of theories address the question of the extent to which performance data can, or should, offer a clear distinction between 'well-performing' and 'poorly performing' hospitals and, in doing so, provide the necessary information to inform subsequent organisational change. The argument is that performance data, expressed as a single, overall indicator, offer an incomplete picture of hospital performance. Therefore, performance indicators alone do not offer an absolute verdict on hospital performance, but instead prompt inquiry and questioning. Carter *et al.*¹⁵⁹ summarise this idea in their depiction of performance data acting as dials, tin openers or alarm bells. They argue that, in an ideal world, performance indicators act like a dial and offer a precise measure of inputs, outputs and outcomes premised on a clear understanding of what good and bad performance entails. However, they recognise that for most public sector organisations there are very few such precise measures, nor is there always absolute agreement as to whether such measures represent 'good' or 'bad' performance. Instead, they suggest that, in most instances, performance indicators act like tin

openers; rather than giving answers, they open up a can of worms and prompt the examination and review of performance. Performance indicators alone thus provide only a partial picture of provider performance.

Similarly, Marshall and Brook¹⁴⁷ argue that:

The utility of comparative data comes less from making absolute judgements about performance than from the discussion arising from using the data to benchmark performance. There is therefore a strong educational component to the effective use of comparative data and resources are required to facilitate this process.

Marshall and Brook¹⁴⁷

Thus, the theory here is that performance data rarely provide a definitive 'answer' regarding the quality of care provided; rather, what leads to change is the discussion and investigation of the underlying cause of the level of performance indicated from the data. Furthermore, such investigations and discussion require resources in order to happen.

The issue of interpretation is a particular challenge for outcomes data. One of the key arguments for the use of PROMs and other outcome data is that they represent the 'ultimate goal' of health care and, as such, health care's success ought to be judged on these data rather than on its activity.⁵ To improve outcomes, it is necessary to change the processes that give rise to those outcomes; however, outcomes data alone do not always provide an indication about the cause of the poor outcomes or what needs to be improved. Thus, there is a tension between the overall objective of performance monitoring and the means to achieve it. Several commentators have criticised the use of outcome measures as performance indicators because they do not distinguish between 'good' and 'poor' care, nor do they provide any pointers for where questioning or investigations might begin to discover the cause of 'poor' performance. For example, Lilford et al.⁷⁰ argue:

The problem that outcome data are poor barometers of clinical quality is viciously confounded by both their inability to discriminate between good and poor performers and the lack of information they convey about how improvements should be made.

Lilford et al.⁷⁰

This criticism has also been levelled at PROMs data; for example, an article in the *Guardian* newspaper about the first release of the 2009–10 PROMs data¹⁶⁰ notes:

PROMs do have their limitations. It is hard, for example, to prove what exactly one hospital is doing different from another that improves outcomes.

Others argue that process measures are better able to signpost the causes of poor care; for example, Mant¹⁶¹ observes:

A second advantage of process measures is that they are easy to interpret. A process measure such as use of aspirin in acute myocardial infarction is a direct measure of quality, whereas hospital-specific mortality from myocardial infarction is only an indirect measure . . . if differences in outcome are observed then alternative explanations need to be considered before one can conclude that the difference reflects true variations in the quality of care. Conversely, a process measure is straightforward to interpret: the more people without contra-indications who receive a proven therapy, the better. A consequence of this is that the necessary remedial action is clearer (use the therapy more often). However, if one does conclude that a higher mortality rate is due to poor quality care, it is not immediately obvious what action needs to be taken, unless perhaps audits of the process of care have been undertaken in parallel.

Reproduced from Mant J, Process vs outcome indicators in the assessment of quality of care. International Journal for Quality in Health Care, 2001, 13(6):p. 475–80, by permission of Oxford University Press¹⁶¹

Here we see a potential way in which outcome data could be made more useful: to conduct an audit of process alongside the collection of outcome indicators, in order to understand why the variation in outcomes has occurred. Others have also noted that the usefulness of measuring process data depends on the strength of the evidence underpinning the process-outcome link. They also note that complex care may not be amenable to single process measures. For example, Long and Fairfield¹⁶² argue that, where there is high-quality evidence that an intervention produces the anticipated outcomes, monitoring process alongside an audit of any adverse events may be sufficient. However, in circumstances where the evidence is mixed or of poor quality, then scrutinising outcomes, together with a 'close description and knowledge of the nature of the process of treatment', is important to establish that the desired outcomes of care have in fact been achieved.

These theories encapsulate the idea that performance data alone do not provide an absolute judgement on the quality of care provided. Instead, performance data may raise questions about the quality of care, which then need to be explored. Furthermore, this process of enquiry is also a process of education, which, in itself, is important to stimulate QI. They also contest different strengths and weaknesses for process and outcome data in the extent to which they may give pointers to this enquiry process. This rests on the strength of the causal link between processes and outcomes.

Focus on outliers or shifting the mean to improve care

Finally, a further debate centres on whether the purpose of performance data feedback is to stimulate improvements by providers classed as 'poor performing outliers' or whether the aim is to prompt improvements in all providers, irrespective of the level of their performance. This clearly has implications for which providers are expected to respond to performance data and how they are expected to respond. Much of the guidance on the national PROMs programme and the presentation of the data using funnel plots is geared towards enabling the identification of outliers (both positive and negative). Although the guidance issued by the Department of Health¹⁰⁹ for identifying potential outliers attempts to give equal emphasis on identifying both positive (i.e. high performing) and negative (i.e. poor performing) outliers, the suggestions for how potential outliers are expected to respond, appear, implicitly at least, to have a subtle focus on stimulating change in poor performing organisations. For example, the summary of key points on page 6 advises that:

The publication of a list of potential outliers will be published as part of or alongside the quarterly PROMs publication. It would be the responsibility of the provider to take to action and explore and improve their performance. We recommend that:

- *The IC's [HSCIC's] participation and response rates table be used by providers to assess the quality of their data. When rates are low, providers would be expected to take action to improve them,*
- *Providers consider if there are other factors which may explain their presented results, other than variation in performance,*
- *Where possible, comparative information be provided to help organisations identified as potential outliers for example, how they compare with other providers on pre-operative scores or on patient characteristics.*

Reproduced from Department of Health under the Open Government Licence v3.0¹⁰⁹

However, other commentators have argued that larger health gains can be achieved if the focus was on supporting improvement in all organisations, not just poor performers. For example, Lilford et al.¹⁵⁰ contend that:

... focus on outliers is only one aspect of improvement. Improving clinical and managerial processes in the remaining organisations can achieve much more health gain by shifting the mean ... the actions required to pull back an outlier as opposed to shifting the mean ... are fundamentally different. The first needs investigation of a probable special cause, whereas the second action needs focus on

improvement by means of a wide array of process-improving tools such as the plan-do-study-act cycle. For these reasons, hospitals should monitor their own data – institutional processes, throughput, clinical processes, outcome (especially trends over time), and above all clinical process.

Reprinted from The Lancet, vol. 363, Lilford R, Mohammed MA, Spiegelhalter D, Thomson R, Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma, pp. 1147–54, © 2004 with permission from Elsevier¹⁵⁰

Thus, Lilford *et al.*¹⁵⁰ point out that if the focus of performance data feedback is to improve the performance of all organisations and not just of poor performers, then all organisations need to use a range of QI techniques, rather than expecting poorly performing organisation to focus on the search for a specific cause of their poor performance.

We noted one further reason for interpreting any data that compare providers with caution discussed in the literature: that of ‘regression to the mean’. There is a statistical likelihood that a provider that is an outlier on one set of data will, in the next tranche of data (e.g. the following year) no longer be an outlier. It will regress to the mean for statistical reasons that may not reflect any true change in performance.¹⁶³ Two potential solutions to avoid this pitfall were voiced in the literature. The first was the idea that it is prudent to monitor providers over several rounds of data collection to detect those that are persistent outliers so as to avoid inappropriately reacting to one set of data. The second was that the actual difference from that expected, not just whether or not a provider is a statistically significant outlier, should also be considered.

Indicator specification

Linked to the issues raised in the previous section, there is debate about whether performance data should be expressed as a continuous variable, or whether targets or thresholds are needed to both interpret the data and incentivise change.¹³⁷ Shaw *et al.*¹³⁷ argue that the use of thresholds or statistical approaches to characterise performance as ‘outlying’ or ‘not outlying’ can lead to arbitrary distinctions between what constitutes ‘good’ and ‘poor’ care. They observe that, within a performance management framework, it is nonsensical to treat performance on one side of this cut-off point as having no cause for concern while considering performance that is not statistically different, but on the other side of the threshold, as problematic.

A further key tension is whether these thresholds should be defined in relation to evidence-based clinical criteria or clinically important changes/cut-off points or in relation to statistical criteria, such as standard deviations above or below the mean population performance. Bird *et al.*¹⁶⁴ make the following comment about national targets based on statistical significance:

... it is unsmart to specify a national target in terms of statistical significance rather than operational terms because, with a sufficiently large denominator, such as from complete enumeration of 600,000 school children’s test results, very small differences of no practical importance achieve statistical significance.

*Bird et al.*¹⁶⁴

In other words, with a large number of events, small differences in performance between units or organisations can be statistically significant without being practically or clinically meaningful.

The community involved in the development of PROMs has recognised the problem that statistically significant differences in PROMs scores following an intervention within a RCT may not reflect clinically important changes. To address this, Jaeschke *et al.*¹⁶⁵ introduced the concept of the ‘minimal clinically important difference’ and developed a methodology to identify it for each PROM. The minimal clinically important difference is defined as:

*The smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management.*¹⁶⁵

This represents a different approach to thresholds based on statistical parameters and instead focuses on changes in PROM scores that would lead a clinician to change the way they manage a patient.

Summary of actionability

The key ideas here are that unless PROMs and other performance data can be understood by recipients, and unless they provide information about the potential cause of any poor performance or offer pointers to how care can be improved, then providers are unlikely to be able to make use of the data to improve patient care. The 'actionability' of data depends on the timeliness of feedback, the way data are presented and the capacity of the organisations and individuals to make sense of and respond to the data. There is some debate about whether or not process data are more actionable than outcome data because they are better able to pinpoint and diagnose where the deficiencies in care lie. Questions are also raised about whether we expect performance data to act like a 'dial' and clearly distinguish between poor and good performers or whether we expect performance data to act like a 'tin opener' and prompt further questioning and investigation and, in doing so, lead to improvements being made. Finally, there is debate about whether the purpose of performance data are to improve the quality of care of all providers, irrespective of their performance, or whether the focus should be on improving the performance of low performers in particular. Linked to this are different ideas about whether such data should be presented as a continuous variable or in relation to a threshold and whether such thresholds should be based on statistical or clinical parameters.

Substantive theories of patient-reported outcome measures feedback

The previous section catalogued the different programme theories underlying the feedback of aggregate PROMs feedback. This shed light on the inner workings of these interventions as perceived by those who design, implement and receive these interventions. In this section, we make connections between these lower-level programme theories and higher-level, more abstract theories, to develop a series of hypotheses that can be tested against empirical studies in our review. By bringing together the programme theories and abstract theories, we can develop a series of more general hypotheses that can be tested against the evidence and produce transferable lessons about how, and in what circumstances, PROMs feedback produces its intended outcomes.

We identified three types of abstract theories underlying the feedback of aggregate PROMs data: audit and feedback, benchmarking and public disclosure.

Audit and feedback

Patient-reported outcome measures feedback is an example of an audit and feedback intervention, as it involves generating a 'summary of clinical performance of health care over a specified period of time aimed at providing information to health professionals to allow them to assess and adjust their performance'.⁹⁰ The idea here is that if providers perceive the health gain produced by surgery in their own hospital to fall below that produced by the average of all hospitals, they will be motivated to narrow the gap by changing their behaviour, that is, changing the way care is provided and thus improving the degree of health gain in their hospital. The assumption here is that clinicians are intrinsically motivated to perform well and that perceiving a gap between their actual performance and the 'ideal' or 'average' may act as a 'wake-up call' to motivate them to change their behaviour in order to improve performance.

We identified several models of audit and feedback, including control theory^{166,167} and feedback intervention theory.¹⁶⁸ They focus largely on explaining how individuals, rather than organisations, might respond to feedback about performance. Both control theory and feedback intervention theory posit that behaviour is goal-driven, and that people change their behaviour in response to feedback about the divergence between their current behaviour and a behavioural goal. Feedback revealing a discrepancy prompts corrective adjustments to behaviour to reduce the discrepancy and proceed towards goal attainment. Control theory hypothesises that if the discrepancy revealed by feedback is too great, or the

feedback recipient lacks skill, motivation or strategies for action, the recipient may disengage from the goal pursuit and 'give up' trying to achieve his or her goal. Similarly, if an individual's performance falls below that of the goal or standard, feedback intervention theory suggests there are four potential responses an individual may adopt: (1) increase their effort to achieve the goal; (2) abandon the standard; (3) change the standard – either lower or higher; and (4) reject the feedback message. These theories suggest that feedback may, therefore, be enhanced through the use of specific performance targets to permit comparison between current and target performance, and action plans to inform behavioural adjustment to reduce discrepancy.

Benchmarking

Patient-reported outcome measures feedback also has a comparative element, such that providers can also compare their own performance with that of other providers in their locality or across England. One of our stakeholders explicitly expressed a hope that 'negative outliers' would identify other local providers who were performing better or who were 'positive outliers' and question: 'what is it that they are doing that we are not?' and 'can we learn anything from them?' This is echoed in the programme theories above, which envisage that providers compare their own performance with that of their peers, and will thus be motivated to respond to be as good as or better than their peers. Those from management and organisational studies would instantly recognise this proposed mechanism as a key component in benchmarking, defined as 'the formal and structured process of searching for those practices which lead to excellent performance, the observation and exchange of information about those practices, the adaptation of those practices to meet the needs of one's own organisation, and their implementation'.¹⁶⁹ The Department of Health defines benchmarking as 'a systematic process in which current practice and care are compared to, and amended to attain, best practice and care'.¹⁷⁰ Thus, benchmarking theories envisage that providers will then identify the practices that lead to this superior performance and implement those practices in their organisation. We identified two theories of benchmarking that have been developed to understand the circumstances under which providers may engage in benchmarking behaviour, the factors that may influence the nature of those activities and the success or otherwise of benchmarking practices in public sector organisations.

Wolfram Cox *et al.*⁹¹ highlight that the benchmarking literature portrays benchmarking as being both a competitive and a collaborative endeavour. Competitive approaches to benchmarking emphasise the idea that benchmarking is a process through which one organisation seeks to surpass the performance of another in order to maintain or increase its market share, while collaborative benchmarking envisages benchmarking as a process of sharing of best practices and mutual learning. This tension relates directly to some of the purported mechanisms through which the feedback and publication of performance data are thought to work contained in the programme theories discussed above, for example through providers fearing a loss to their market share or wishing to be as good as or better than their peers.

Wolfram Cox *et al.*'s⁹¹ theoretical model sought to identify the contingencies that predict the degree of collaboration or competition that will occur in a benchmarking exercise, which in turn will influence the markers of success and the longevity of the benchmarking exercise. Of particular relevance to our review is their proposition that whether benchmarking is collaborative or competitive depends on the benchmarking context. The authors highlight a number of contextual factors that they separate into structural factors, including the degree of geographical separation; the number of partners involved; and dynamic factors, such as who initiates the benchmarking, the primary motivation for initiating the benchmarking and the quality of individual associations. They hypothesise that the larger the geographical distance between benchmarking partners, the less competition between them and, therefore, the greater likelihood that benchmarking will be collaborative. They predict that benchmarking is also more likely to be collaborative when there are a larger number of partners. They also theorise that existing relationships can amplify the nature of benchmarking, so if there is a history of competition between two organisations, benchmarking is more likely to be competitive. They also argue that the longer benchmarking continues and the more frequently the parties meet, they are more likely to be collaborative, because strained relationships dissolve over time and trust develops.

van Helden and Tillema's⁹³ benchmarking theory (*Figure 10*) draws on economic, institutional reasoning and resource dependency theories to develop a model that seeks to explain the potential outcomes of imposed benchmarking in terms of performance improvement and specify the contingencies that might influence the ways in which organisations respond to imposed benchmarking in the public sector. They draw on economic theories to predict the potential outcomes of benchmarking. From an economic perspective, public sector benchmarking is an alternative to market forces and an attempt to stimulate competitive behaviour in order to improve performance. Market forces require that consumers switch from one provider to another, but this mobility is limited in many public services, including the NHS. However, funding bodies and consumers can exercise their power to ensure that their interests are taken into account, by subjecting organisations to increased scrutiny, placing them under direct supervision or contracting out 'failing' services to alternative providers. This echoes the ideas contained in the public accountability programme theories discussed above. Thus, organisations are motivated to improve their performance to secure the support of funding organisations, rather than from fear of losing market share. In turn, benchmarking is expected to improve the performance of all organisations, to reduce variation in performance and to provide a stronger incentive for poor performers.

From institutional reasoning and resource dependency theories, van Helden and Tillema⁹³ predict that an organisation's response to benchmarking is determined by the different types of institutional pressures exerted on it by government, professional groups, interest groups and the general public. They draw on Oliver's¹⁷¹ typology of response patterns, which vary from passive compliance to proactive manipulation of pressures. These response patterns relate directly to some of the unintended consequences reflected in the programme theories discussed above (see *Unintended or adverse consequences of performance data*), such as the gaming of data. van Helden and Tillema⁹³ argue that an organisation's response pattern depends on its willingness and ability to conform to institutional pressures, which, in turn, depends on the reasons the stakeholders have for exerting the pressure (cause), which stakeholders exert the pressure (constituents), what the pressures are (content), the way the pressure is exerted (control) and environment in which the pressure is exerted (context).

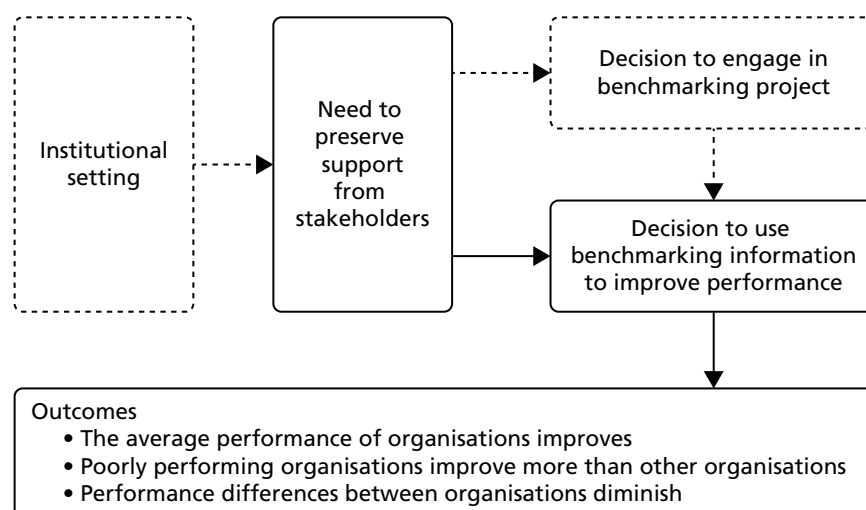


FIGURE 10 van Helden and Tillema's benchmarking model. Reproduced from van Helden and Tillema⁹³ with permission from John Wiley and Sons.

Of particular relevance to this review is van Helden and Tilemma's⁹³ hypothesis that benchmarking provides a stronger incentive to improve performance for poorly performing organisations, as these organisations are under most pressure from funding bodies. They also theorise that the lower a public sector organisation's acceptance of poor benchmarking scores, the more likely it is that it will not improve performance. This relates directly to the programme theory discussed above, that providers will not take steps to improve patient care unless performance data are perceived as accurate and credible. The authors also hypothesise that the more an indicator can be regarded as a 'soft indicator', the more likely it is that – in reaction to a low score on this indicator – a public sector organisation will improve the presentation of the indicator rather than the performance itself. This suggests that gaming is more likely to occur where providers are under pressure from funding bodies to improve but the perceived credibility of the data is low.

Public disclosure

Patient-reported outcome measures data are also a form of public disclosure of performance data, or, as some have less delicately put it, a form of 'naming and shaming'.²¹ This directly links to the programme theories that providers will be motivated to respond from fear of losing their market share (as patients choose high-performing hospitals) through concerns about damage to their reputation. Pawson²¹ conducted a realist synthesis of the relative fortunes of different public disclosure interventions in different policy sectors and, subsequently, Exworthy *et al.*¹⁷² considered how this underlying programme theory can be applied to understand the potential impact of the public disclosure of clinical performance in cardiac surgery. We can use some of the theoretical propositions in this model to examine the contexts and mechanisms through which public disclosure of hospital performance data may lead to improvements in patient care.

Pawson's²¹ theory identifies four stages to public disclosure interventions:

1. identification, in which performance is classified, measured or rated
2. naming, in which performance is disclosed, disseminated or publicised
3. public sanction, in which 'responsible bodies' (e.g. Clinical Commissioning Groups, patients, the CQC) respond by reprimanding, censuring, influencing, supervising or controlling poorly performing providers (e.g. 'negative outliers')
4. recipient response, in which poorly performing providers respond by changing their behaviour – ideally, by improving their performance.

In our review, 'identification' corresponds most directly to the ways in which performance data are produced, and relates to our programme theories on data credibility. 'Naming' relates most clearly to the ways in which performance data are both presented and publicised, in particular the role of the media in this process. 'Public sanction' reflects the ways in which those who may directly influence providers, such as patients, GPs and Clinical Commissioning Groups but also regulatory bodies such as the CQC and the General Medical Council, respond. 'Recipient response' is how providers then choose to respond.

The theory also identifies a number of potential unintended consequences to public disclosure. Of particular relevance to our review is the idea of culprit misidentification, where performance is classified inappropriately, with the measure being over- or underdiscriminating or lacking in 'risk' adjustments. This relates directly to problems with case-mix adjustment of performance data identified in our programme theories. Pawson's theory suggests that culprit misidentification is more likely to happen when the behaviour in question is complex. A further problem is mismanaged dissemination, where disclosure is poorly managed by sparse or excessive publicity, over-restricted or overstretched targeting, overcomplexity or oversimplification in presentation or wrangles about the meaning of the information. This relates to concerns about data interpretability and about the role of the media in either 'translating' or 'misrepresenting', as identified in our programme theories. Pawson also identifies a number of unintended outcomes that share similarities with van Helden and Tilemma's⁹³ benchmarking theory, above, and our programme theories, such as when

individuals or institutions react to public disclosure by accepting the label and amplifying poor performance or reinterpreting the label or adopting a perverse modification to behaviour. Pawson also suggests that public disclosure may improve provider performance only in conjunction with other sanctions or incentives, which suggests the importance of exploring how public disclosure of performance may work in the context of financial incentives and sanctions.

An overall programme theory to guide the review

After discussion with stakeholders and within the project team, we agreed to focus our review on understanding the contexts in which, and mechanisms through which, providers respond to performance data to improve patient care. Drawing on the above programme and abstract theories, we developed an overall programme theory to structure our review (*Figure 11*).

This programme theory highlights a range of contextual factors identified by our programme and abstract theories that may influence whether or not providers respond to performance data; these include whether or not the data are perceived as accurate and credible, whether or not the data are timely, whether or not providers are able to interpret the data, and whether or not the data provide information on the likely or possible causes of poor care and wider contextual factors, such as the use of financial incentives or sanctions. The mechanisms through which providers may respond to performance data are also specified, including a wish to protect their market share, a wish to protect their professional reputation, a wish to be as good as or better than their peers and an intrinsic desire to provide high-quality care. The possible outcomes include taking steps to improve care, and unintended or adverse outcomes such as ignoring or dismissing the data, gaming and effort substitution. This programme theory will act as a framework for the evidence review, in which the connections between these contexts, mechanisms and outcomes can be tested and refined to provide an explanation of the process through which providers respond to performance data.

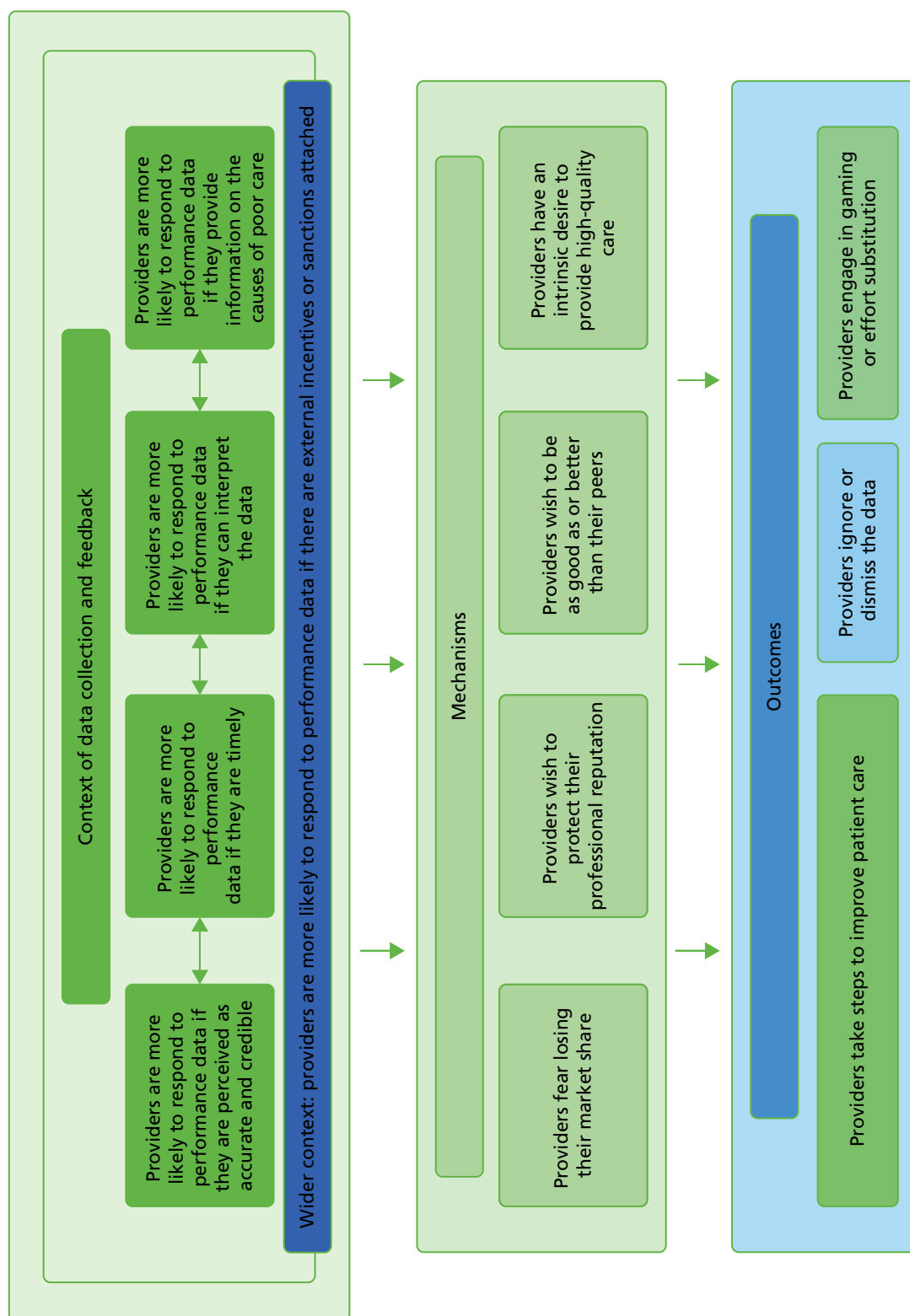


FIGURE 11 An overall programme theory to guide the review.

Chapter 4 Feedback of aggregate patient-reported outcome measures and performance data: reviewing mechanisms

Introduction

In this chapter, we review evidence to test and revise our theories of the mechanisms through which the feedback and public reporting of hospital quality improves patient care. Where possible, we focus on studies that have evaluated the use and impact of PROMs and PREMs as indicators of the quality of patient care to test our theories. However, for many of our theories, both in this chapter and in *Chapter 5*, there were few studies of PROMs or PREMs specifically that were relevant to our theories. Therefore, we draw on a wide range of evidence, including studies examining the impact of, and providers' responses to, other forms of performance data. We do so because these interventions share the same underlying programme theories as those reviewed in *Chapter 3*. In realist synthesis, the unit of analysis is the programme theory, not the intervention itself. Therefore, as cardiac report cards and other forms of hospital report cards and performance data share the same underlying programme theories as the feedback and public report of PROMs data, we consider these studies together. Here we briefly outline the overall structure of the chapter before beginning our synthesis.

At its most basic, the underlying programme theory (theory 1) is that the feedback of PROMs and other performance data will lead to improvements in patient care and, ultimately, to better patient outcomes. We begin our synthesis by testing this unrefined theory by examining studies and systematic reviews that examine whether or not the feedback of PROMs and other performance data leads to improved outcomes for patients. We then move on to test theory 2, that the medium of feedback influences provider behaviour according to whether feedback is delivered privately or publicly. Following this, we test a number of theories which present different ideas about how the feedback and public reporting of provider performance data is intended to work. These theories are:

- Theory 3: the public reporting of poor performance threatens the provider's market share and provokes a significant provider response.
- Theory 4: providers perceive that report cards damage their professional or their hospital's reputation.
- Theory 5: providers respond to the feedback and public reporting of performance data by comparing themselves with their peers, as envisaged in abstract theories of benchmarking.

Finally, we consider how patients are expected to respond to performance data and we test theory 6, that patients choose hospitals on the basis of public reports of hospital quality.

Theory 1: the feedback of patient-reported outcome measures or performance data leads to improved patient care

In *Chapter 3*, we reviewed the different programme theories underlying the feedback of PROMs and other performance data to improve patient care. At its most basic, the underlying programme theory is that feedback of these data will lead to improvements in patient care and, ultimately, to better patient outcomes. We begin our synthesis by testing this unrefined theory through briefly reviewing the literature that explores whether or not the feedback of PROMs and other performance data leads to improved outcomes for patients. We first consider the national PROMs programme in England, which has a relatively short history. Unsurprisingly, we found few studies that had evaluated whether or not the programme's introduction had led to improvements in patient outcomes following surgery for hip and knee replacement,

hernia repair and varicose vein procedures in England. The feedback and public disclosure of performance data has had a long and somewhat turbulent history. There are a large number of studies that have evaluated whether or not this programme has led to improvements in patient care and outcomes. Much of this evidence comes from the USA and Canada, where public reporting was first introduced in the early 1990s. These studies have been synthesised in the form of traditional narrative systematic reviews, which attempt to explore whether or not public reporting of performance data improves patient outcomes and the process of care. These studies thus provide an introduction to, and an indication of, the broad contours of this literature. They provide us with an understanding of the outcome patterns which reflect answers to the question of whether or not these programmes improve the quality of patient care and patient outcomes, but they do not attempt to explain why, how or when these outcomes do (or do not) come about.

Theory 1a: patient-reported outcome measures feedback will lead to improved patient outcomes

Here we review two studies that examined whether or not the feedback of aggregate PROMs data led to improvements in patient outcomes.

Varagunam *et al.*¹⁷³

Varagunam *et al.*¹⁷³ evaluated the impact of the national PROMs programme in England, and tested whether or not the programme delivered on its policy aims of changing the selection of patients for surgery, reducing the variation among providers in the degree of health gain achieved by patients and improving postoperative outcomes for patients. Using the national PROMs data set, the authors analysed changes in mean preoperative scores and mean adjusted postoperative scores using standardised effect sizes on both disease-specific and generic PROMs for each procedure between April 2009 and March 2012. They found that preoperative severity increased slightly for varicose vein procedures and, to a lesser extent, hip surgery, but not for knee or groin hernia surgery, suggesting that the programme did not have a significant impact on the ways in which patients were selected for surgery. They found little variation between providers and that this variation did not change over time. There was only a modest consistency in providers labelled as outliers over time for hip and knee replacement and much less consistency for the other two procedures. The authors attribute this in part to regression to the mean. They also found slight improvements in health gain for hip and knee replacements but, while these changes were statistically significant, they could not be considered clinically significant. They found no changes over time in health gain for groin hernia repair and a slight worsening in the degree of health gain for varicose vein surgery.

Overall, the authors conclude that the national PROMs programme had 'little impact' on preoperative severity and outcomes for patients and suggest that more time is needed to fully evaluate the impact of this programme. In their discussion, the authors attribute this finding to two factors. First, the delay between data collection and feedback of PROMs data meant that providers only received the first release of the finalised data on outcomes in April 2011 and, as such, these data were unlikely to have had any impact on provider practices, and thus outcomes, until the fourth year of the programme. Second, until 2011, the data were fed back as a large spreadsheet, which the authors argue was difficult for clinicians, managers and patients to understand. The use of funnel plots, which enabled providers to compare their own performance with that of others, only commenced in 2011. In terms of the theories under test, both the lack of timeliness of the data and the difficulty in interpreting them may have limited the ability of providers both to compare themselves with others and to identify any gaps between their own performance and that of others. Subsequently, the authors have also suggested that the lack of impact may be due to a ceiling effect; there is little variation between providers, with the vast majority delivering good levels of health gain in clinical terms, and, as such, there is little room for improvement in performance for these procedures.

Boyce and Browne¹⁷⁴

This study aimed to examine whether or not providing surgeons with peer-benchmarked feedback about the patient-reported outcomes for their patients improved the outcomes of their future patients. The study was conducted in Ireland, where there is no national programme to feed back hospital-level PROMs data, as exists in England. Furthermore, the feedback was provided at the level of the surgeon, rather than of

the hospital, and was privately rather than publicly reported. The authors used a cluster RCT design, with surgeon as the cluster unit.

The trial focused on high-volume orthopaedic surgeons, defined as surgeons who were responsible for at least 100 primary hip operations during the previous year. A total of 21 surgeons were recruited to the trial: 10 were randomised to the control group and 11 were randomised to the intervention group. Intervention group surgeons received a case-mix-adjusted peer-benchmarked comparison of their patients' mean change in the OHS in the form of a 'caterpillar chart'. This plotted the patients' mean change in the OHS and confidence intervals for each of the anonymised 21 surgeons, ordered from highest to lowest, with the recipient's own performance clearly marked. Intervention surgeons also received feedback on the proportion of their patients who reported an overall improvement in their hip problem and the proportion who reported having at least one of four postoperative complications. Data for this feedback were collected during a period of 11 months prior to the provision of the feedback. The surgeons also received a 9-minute educational video about the interpretation of PROMs data. The primary outcome for the trial was the difference in the mean postoperative OHS between patients in the intervention and control arms who were operated on during a period of 11 months following the provision of feedback. The secondary outcomes included the Hip Osteoarthritis Outcome Score and the EQ-5D.

Oxford Hip Score data were analysed using a linear mixed-effects regression model, adjusting for differences in patient characteristics. The authors found no statistically significant differences in mean postoperative OHS, Hip Osteoarthritis Outcome Score or EQ-5D. They also found no statistically significant differences in the proportion of patients who experienced a complication between the intervention and control arms. The authors conclude that 'embedding PROMs within quality assurance and improvement frameworks is unlikely to lead to patient benefit'.¹⁷⁴ The authors attribute the lack of impact to a number of issues, including surgeons' scepticism about the value of the data, the burden of data collection and the lack of sensitivity and specificity of PROMs as indicators of clinical performance. They also note in their conclusions that 'Performance monitoring can provide information about whether healthcare professionals perform better or worse than their peers, but does not explain why performance differs'.¹⁷⁴ This echoes the theory identified in *Chapter 3*, that outcome measures do not provide information on which aspects of process are attributable to poor performance. The authors go on to argue that in order to identify the source of any problems, clinicians will have to undertake additional audit activities, which 'assumes that professionals have the time, resources, knowledge, expertise, flexibility and willingness to implement such activities'.¹⁷⁴

Theory 1b: the public reporting of performance data leads to improved patient outcomes

Several systematic reviews have collated the evidence from quantitative studies examining the impact of the public reporting of performance data on patient outcomes and the process of care.^{19,26,27} Here, we summarise the findings from the most recent review conducted by Totten *et al.*²⁷ for the Agency for Health Research and Quality in the USA.

Totten *et al.*²⁷

This was a large systematic review of over 198 studies to examine whether or not the public reporting of performance resulted in improvement to quality of health care or health-care delivery structures and processes. These studies largely focused on evaluating reporting systems implemented in the USA and Canada, both of which have a longer history of public reporting initiatives than the UK. As such, the transferability of the findings from this review to the UK context will be limited by differences in the funding of the UK, US and Canadian health systems and will also vary depending on the characteristics of the individual reporting system under study. The majority of the studies evaluated the public reporting of mortality rates, although some programmes also included other outcome, process and structure indicators. The precise nature of these programmes varied enormously, but the review was not designed to provide a detailed exploration of the ways in which these variations might have influenced the outcomes. Although the review aimed to be comprehensive and used systematic search strategies, the authors

acknowledged that they could not rule out the possibility that some studies might have been missed. The main analyses in the review focused on quantitative studies. The authors provided a descriptive summary of qualitative studies and their findings, but made but no attempt to synthesise this evidence.

Across all the different reporting systems studied, overall, the review found a small decline in mortality following public reporting after controlling for trends in a reduction of mortality; however, individual studies varied in their findings. For example, studies examining the impact of cardiac public reporting programmes on mortality rates found a variable picture: eight studies found a decrease in mortality rates over time,^{29–36} whereas another four studies^{37–40} found no changes in mortality rates over time. Similarly, although most studies examining the impact of public reporting on process indicators found an improvement in hospital quality, this varied from a ‘slight’ improvement to a ‘significant’ improvement in quality. The review also examined studies that specifically attempted to understand the contextual factors that may influence the impact of public reporting and concluded that QI is more likely to occur in a competitive market and following low performance.

Theory 1 summary

In summary, only one study has examined whether or not the national PROMs programme in England led to improved patient outcomes, and it found a lack of impact.¹⁷³ The study itself was not designed to investigate why the national PROMs programme has, so far, not led to improvements in patient outcomes. The authors hypothesised that this was largely due to the infancy of the programme itself and, in line with some of the programme theories reviewed in *Chapter 3*, because the data were not timely and were difficult to interpret. They also suggested that the PROMs programme targeted procedures for which performance in general was already good and, as such, there was little room for further improvement. Another study examined the provision of peer-benchmarked surgeon-level feedback to surgeons and found no impact on patient outcomes.¹⁷⁴ The authors argued that PROMs feedback does not provide information on why performance differs and assumes that clinicians have the skills and resources to undertake additional investigations to uncover the cause of any poor performance.

There is a much larger number of studies examining the impact of public reporting systems in the USA and Canada on patient outcomes and patient care. The most recent review of this literature found variation in whether or not public reporting leads to improved patient outcomes and patient care.²⁷ The review pointed to some possible contextual factors that may explain this: that improvements in patient care are greater when there is greater market competition and when providers are low performers. However, this review was not designed to examine the processes through which or contexts in which the feedback of performance data improves patient care.

In subsequent sections of our realist synthesis we attempt to delve into the ‘black box’ to understand the contexts within which, and processes through which, the feedback of performance data may lead to improvement in patient care. We do this by testing the programme theories set out in *Chapter 3*. In this chapter, we test and refine theories about the mechanisms through which the feedback and public reporting of performance data is thought to work. In *Chapter 5*, we review theories about the contextual factors that support or constrain these mechanisms.

Reviewing mechanisms: how and why does performance feedback lead to improvements in patient care?

A key assumption underlying the feedback of performance data is that receiving this feedback will stimulate providers to make improvements to patient care. One of the tasks of realist synthesis is to adjudicate between rival programme theories, or between competing mechanisms about how a programme works.⁸² The goal here is not to produce ‘winners and losers’ but to understand and explain the circumstances in which different mechanisms are triggered. Through an iterative process of testing theories in relation to the empirical literature, we can refine these theories and thus improve our understanding of how and why programmes work. As we discussed in *Chapter 3*, there is a plethora of different theories about the mechanisms through which the feedback and public disclosure of performance may (or may not) stimulate

providers to take steps to improve patient care. The goal of this section of our synthesis is to put these theories to the test. We begin by briefly revisiting these theories.

Hibbard *et al.*¹¹⁸ hypothesised three different mechanisms through which the feedback and public reporting of performance might stimulate providers to embark on QI activities, which we have adapted to our synthesis. Providers may respond because of one of the following.

- ‘Intrinsic motivation’: their professional ethos means they are intrinsically motivated to maintain good patient care and will take steps to improve if feedback highlights there is a gap between their performance and expected standards of patient care.
- ‘Market share’: they feel threatened by the potential loss of market share that could occur if patients decided to choose alternative, higher-performing providers.
- ‘Professional reputation’: they wish to protect their professional or institutional reputation, which may have been damaged by being labelled a poor performer in public.

The first mechanism resonates with audit and feedback theories that simply highlighting a gap between provider performance and expected standards will prompt providers to respond. It may occur as a result of public reporting but does not *rely* on the public reporting of performance to occur. The assumption is that the realisation of a gap between their own performance and the target or ideal performance will be sufficient motivation to prompt individuals or organisations to improve, irrespective of whether this feedback occurs privately or publicly. As such, we would expect providers to take steps to improve care even if they receive performance feedback privately.

The second and third mechanisms reflect public disclosure theories and are stimulated through the process of *public* rather than private reporting. The first assumes that when performance data are made available to them, the public will choose higher-quality providers and, in turn, providers will take steps to improve care either to *prevent* shifts in their market share or in response to actual changes in their market share. We recognise that fear of losing market share is a much stronger mechanism in USA and Canadian health systems. These systems are based on health insurance rather than government funding raised through taxation. In these systems, providers compete for business from health insurers, employers and patients. However, in England, successive government reforms, starting with the purchaser–provider split in 1989, have attempted to introduce greater competition into the NHS, with the aim of driving up quality.¹⁷⁵

The second relates to the ‘shaming’ mechanism in public reporting: it assumes that public reporting may damage the reputation of either an individual clinician or an organisation and that they act to improve care in order to restore or protect this reputation. Reputational damage is assumed to occur when the public, peers or other stakeholders have a *worse* opinion about the performance of that individual or organisation than they did before they were exposed to a report about the performance of a particular individual or organisation. Both rely on the *public* (rather than private) disclosure of performance to occur, as it is the real (or imagined) responses of ‘responsible’ bodies or stakeholders that in turn stimulate a response in those whose performance is being publicly disclosed.

The ideas and assumptions about the mechanisms through which the feedback of PROMs data stimulates providers to improve care also drew on benchmarking theories.^{91,93} These theories assume that providers will be motivated to respond because of:

- ‘Competitive benchmarking’: they are competitive and wish to be as good as or better than their peers.
- ‘Collaborative benchmarking’: they improve patient care through learning about and implementing the best practices of ‘high-performing’ organisations as a result of the sharing of information.

The motivation to outperform or be as good as peers is the central mechanism motivating organisations to take steps to improve performance within theories of competitive benchmarking.⁹¹ These assume that individuals or organisations wish to outperform each other. To work, it relies on the provision of information that enables organisations or individuals to compare their own performance with that of

others. This may or may not be publicly reported. In contrast, theories of collaborative benchmarking focus on the role of learning from the best practices of others as the mechanism through which organisations improve their performance. Here, the process is assumed to be one of mutual benefit and collaboration. Furthermore, benchmarking theories derived from institutional reasoning and resource dependency theories^{93,94} hypothesise that low-performing organisations will feel a greater pressure to respond to poor performance because they are most under threat from sanctions (such a loss of funding) issued by stakeholders than high performers.

We now put these rival theories to the test through exposing each theory or set of competing theories to the empirical literature.

Theory 2: feedback influences provider behaviour according to whether it is delivered privately (confidentially) or publicly

As outlined above, if providers are motivated by an intrinsic desire to improve care, they would be motivated to respond to feedback on performance that was provided privately. It is assumed that the realisation of a gap between their own performance and the average or 'ideal' performance is sufficient to motivate providers to take steps to improve care, without the need for this gap to be made public or shared with others. We can test this theory by reviewing studies that have examined the impact of confidential feedback of performance to providers such as feedback from medical registries and clinical audits. However, others have argued that the public reporting of performance data places increased pressure on providers to respond over and above private reporting. Therefore, to test the theory that the medium of reporting influences providers' responses, we can examine studies that have compared the impact of public versus private reporting of hospital performance on patient outcomes and QI efforts. If providers have an intrinsic desire to improve, it would be expected that private reporting would be as effective as public reporting in bringing about improvements in care. However, if public reporting does place increased pressure on providers to respond, we would expect to see that public reporting has a greater impact on QI efforts or patient outcomes than private reporting. We start by reviewing studies that have examined the impact of confidential feedback to providers in the form of medical registries and clinical audits. We then review studies that have compared the impact of private versus public reporting of performance.

Theory 2a: providers are motivated to respond to private/confidential feedback on performance

van der Veer et al.¹⁷⁶

This systematic review examined whether or not feedback from medical registries had an impact on the quality of patient care, as well as charting the details of how these data were fed back and the factors noted as barriers and success factors by the authors of the studies reviewed. The authors included both randomised and non-randomised studies that had explored whether or not feedback impacted on the process and outcomes of patient care. They included only studies in which feedback was provided to health-care professionals or the departments in which they worked, rather than to patients or policy-makers. Therefore, this review enables us to test the theory that the private feedback of performance improves the process or outcomes of patient care.

A total of 53 papers reporting on 47 different registries were included in the review. The authors highlighted that the ways in which feedback from medical registries was fed back to providers was heterogeneous in terms of the main purpose of the registry, the medium through which feedback was provided, the specificity of the feedback, the benchmark used to compare provider performance, the frequency with which feedback was given and the time lag between data collection and feedback. As such, the authors found it difficult to 'make straightforward comparisons between feedback initiatives and to draw definite conclusions on the effectiveness of feedback'. Of the 53 included papers, the authors report that 22 studies (reported in 24 papers) evaluated the effect of feedback on one or more primary

clinical outcome or process measure. The authors provide little detail about the different process and outcome measures used in the studies they reviewed.

The authors reported that, of these 22 papers, four found a positive effect on all measures, eight found a mix of positive and no effect and 10 did not find any effect. Across the 22 studies, the impact of private feedback was evaluated on 43 process measures and 36 outcome measures, although it is not clear how these measures were distributed among the 22 studies. A total of 26 out of the 43 process measures and 5 out of the 36 outcome measures were positively affected by feedback. This may reflect a provider's preference for process measures because they offer clearer guidance on actions needed to stimulate change. However, it may indicate that it is simply more difficult to show an impact of any changes on outcome measures because they do not have a one-to-one relationship with process-of-care measures but rely on a multifactorial set of processes. The authors also conducted one subgroup analysis to examine the effect of feedback for those registries in which QI was not the primary goal of the medical registry but was developed for other reasons ($n = 9$ studies). Of these nine studies, two found a positive effect, eight found a mix of positive and negative effects and 10 did not find any effect.

The authors also recorded the perceived barriers to and facilitators of effective feedback. It is not clear from their analysis whether exploring the barriers and facilitators was the primary goal of the studies reviewed, or whether these were the opinions of the authors of the reviewed studies, based on their experiences during the study. They note that a lack of trust in data quality was the most common issue raised with the studies, while the most common success criteria were the timeliness of the data and trust in data quality.

This was not a high-quality systematic review; it is not clear how the authors assessed the quality of the studies included in their review, as they provided little detail on the primary outcomes measured in the studies reviewed and their analysis of the effect of feedback on outcomes also lacked clarity. However, their recording of the heterogeneity of the feedback characteristics was detailed. Despite this review's flaws, we can draw a number of conclusions from it. The first is that medical registries vary in their primary purpose and in how the feedback on performance is fed back to providers. Variation in the quality of the studies apart, it is, therefore, not surprising that the impact of feedback from medical registries was also heterogeneous. The review also suggested two possible theories that could be explored further: that provider trust in these data may be an important determinant of their impact and that process-of-care measures are preferred by providers as a target for change and/or better reflect changes made to patient care.

Taylor *et al.*¹⁷⁷

The authors conducted a survey of clinical audit leads responsible for four national clinical audits across NHS trusts in the UK, followed by interviews with a convenience subsample of respondents. The audits chosen for examination were:

- National Oesophago-Gastric Cancer Audit
- National Bowel Cancer Audit
- National Head and Neck Cancer Audit
- National Lung Cancer Audit.

The survey explored respondents' perceptions of the purpose of, impact of and barriers to using feedback. The interviews further probed the findings of the survey and explored what lay behind respondents' responses. Of 607 audit leads contacted, 274 (45%) completed a questionnaire, with response rates being higher for the National Oesophago-Gastric Cancer Audit and the National Bowel Cancer Audit. From 274 questionnaire respondents, 32 volunteered to be interviewed. Both surveys and interviews may be subject to recall and respondent bias; it may be that those who responded to the survey and interviews were those who had more favourable opinions of audit. However, the study provides useful evidence to test the theory that providers are motivated to respond to private feedback.

The findings from the survey suggested that the majority of respondents perceived the audits as useful to identify opportunities for QI (88%), to facilitate team discussions on quality and safety issues (86%) and to benchmark outcomes at their own NHS trust with those of their peers (84%). Roughly two-thirds of respondents indicated that the audits had increased their awareness of levels of performance and practice patterns among their peers. These findings did not differ among the different audits. These findings support theories of audit and feedback and benchmarking, that the feedback increases awareness of peer performance and enables recipients to identify opportunities for improvement. Just over half of all respondents (56%) indicated that they had implemented service improvements and 42% indicated they had changed aspects of their own clinical practice. Respondents from the National Lung Cancer Audit were significantly more likely to report implementing service improvements and changing practice. This might have been because these audit teams were also supported by the Improving Lung Cancer Outcomes Project,¹⁷⁸ which ran between 2010 and 2012 and supported multidisciplinary teams to come together, identify practice variation and share best practice. The other audit teams did not receive such additional support.

The interviews suggested that the audits were perceived as an authoritative snapshot of current clinical practice, owing to the credibility of the professional societies associated with the audits. One interviewee stated: 'I don't think DAHNO [National Head and Neck Cancer Audit] has ever told us anything we didn't already know, it's just given us data that's authoritative . . . it's providing data that makes people realise we're telling the truth.'¹⁷⁷ This credibility was perceived as important when making a case to hospital managers for investment to implement changes. The interviews also explored the barriers to the use of audit data to make QI. A frequently cited issue was that the quality of these data, in terms of both the methods used and the resources available to collect the data, often limited their use. One interviewee highlighted that audit data had to compete for priority with national targets and, as such, fewer resources were devoted to ensuring audit data were collected accurately: 'The trust . . . doesn't appear to recognise the need for quality of clinical data to be collected accurately, because . . . it isn't a national target that they are immediately judged on.'¹⁷⁷ The second key issue was the availability and appropriate presentation of locally benchmarked data to enable trusts to compare themselves with local NHS trusts. The authors note that such comparative data 'motivated trusts to improve, driven on by a competitive spirit'.¹⁷⁷ The authors observed that those interviewees who had been able to make changes had done so in the light of looking at how they compared with trusts locally, while those who did not make changes did not have these data available to them. These findings support theories of benchmarking, that peer comparison motivates providers to improve through peer competition.

Theory 2b: public reporting of performance places additional pressure on providers to respond

Guru et al.³⁹

In Ontario, Canada, coronary artery bypass graft surgery performance report cards, measuring risk-adjusted in-hospital mortality, were first introduced in 1993. Between 1993 and 1999, the report cards were shared confidentially with hospitals, with the aim of motivating QI activities. From 1999 onwards, with the agreement of the cardiac surgeons in the province, the report cards were made publicly available. This 'natural experiment' enabled Guru et al.³⁹ to examine the additional impact of public reporting over and above the impact of private reporting, by examining changes in mortality rates during this time. To do this, the authors compared changes in risk-adjusted rates of in-hospital 30-day mortality in Ontario during the period of private reporting (1993–9) with those during the period of public reporting (1999–2001). They also compared risk-adjusted in-hospital 30-day mortality rates in Ontario with those in the rest of Canada to assess whether or not the changes in rates in Ontario were different from those in rest of Canada, where neither private nor public reporting took place. The authors also examined changes in all-cause mortality and length of stay in Ontario during the same time period, indicators that were fed back to surgeons in Ontario during the observed time period. This allowed a comparison between changes in indicators that were and were not fed back to clinicians.

The authors found a steady decline in risk-adjusted 30-day mortality rates following the introduction of private reporting, with the greatest decrease occurring in 1994 (29% reduction), immediately after private reporting began. The introduction of public reporting appeared to have no additional impact on decreasing mortality rates, and there was a relative increase of 2% in 30-day mortality rates immediately following the introduction of public reporting. There was also a decline in 30-day mortality rates across the rest of Canada during the same time period, but this took longer to reach the low rates seen in Ontario. There were no clinically significant changes in 30-day all-cause mortality or readmission rates in Ontario during the same time period. In their discussion, the authors note that, for hospitals exhibiting poor performance, hospital managers requested reports at the level of the surgeon or undertook additional data collection to establish the potential cause of the problems and took action to deal with any issues identified. They also acknowledged that they could not rule out the possibility that private reporting might have simply accelerated improvements that would have occurred anyway, or that, had public reporting been introduced at the same time as private reporting, the impact on mortality rates might have been larger or have occurred more quickly.

In terms of the theories under test, the authors argued that these findings support the hypothesis that private reporting of performance to providers stimulates QI efforts and that public reporting has no additional impact on efforts to improve quality. As such, the additional pressure created by public reporting had no impact on the providers' responses to such data. However, it is difficult to separate the impact private feedback or public reporting from general trend in the improvement in mortality rates following cardiac surgery across all countries during the time period. It is not clear that these improvements can be solely attributed to either private or public reporting of performance, as there were both changes and improvements in the techniques for performing cardiac surgery occurring at the same time.

Bridgewater *et al.*³⁵

This study was a retrospective analysis of prospectively collected data on the outcomes of cardiac surgery across four centres in the north-west of England. The authors sought to explore whether or not there had been changes in the observed, predicted and risk-adjusted mortality rates following cardiac surgery since these data were publicly reported. The time period observed was from April 1997 to March 2005. The authors sought to test the hypothesis that public reporting had led to improvements in mortality rates by comparing rates before and after the introduction of public reporting. Thus, in this study, the decision about the cut-off point between private reporting and public reporting is a crucial aspect of its validity. The authors argued that the public reporting of surgical mortality rates for cardiac surgery in the UK first began in 2001, when an independent organisation, Dr Foster, published unadjusted named cardiac surgery mortality rates for all hospitals performing such surgery in the UK. Therefore, the authors subdivided the observed time period into 'pre public disclosure' (or private feedback) (April 1997–March 2001) and 'post public disclosure' (April 2001–March 2005).

The authors found a significant fall in observed mortality from 2.4% in 1997–8 to 1.8% in 2004–5 ($p = 0.014$). There was also a significant reduction in the observed-to-expected mortality ratio from 0.8 in 1997–8 to 0.51 in 2004–5 ($p < 0.05$). Both observed and risk-adjusted mortality rates were significantly lower in the post-public disclosure period than in the pre-public disclosure period. The authors are careful in their discussion not to directly attribute these findings to the introduction of public disclosure, instead merely noting that these findings were observed 'since the introduction of public disclosure'.³⁵ However, in a later paragraph, they did acknowledge that their intent was to study the 'effects of publication of results'.³⁵

The authors acknowledged several limitations to their study. They accepted that, over the study period, an increasing number of patients were treated with percutaneous coronary intervention, and improvements in percutaneous coronary intervention techniques also occurred. Therefore, it is unlikely that improvements in mortality rates were solely due to the introduction of public reporting. Furthermore, the introduction of public reporting may not have equated with providers feeling accountable. Although the Dr Foster report was published in 2001, Bridgewater¹²⁷ noted, in a different journal article, that this publication was roundly condemned by surgeons at the time, owing to its lack of case-mix adjustment, and it could be argued that it did not result in surgeons feeling accountable or in attempts to improve their practice. In this paper, the

authors argued that 'it became clear to most members of the surgical community that publication of their outcomes was inevitable between 2001 and 2005' and thus 'it is reasonable to use this date as a cut off point'.³⁵ However, they acknowledged that there is not a clearly defined date that demarcates the introduction of public accountability. Finally, they also noted that public reporting was introduced following many years of the provision of structured, private feedback to doctors, which was also likely to have improved surgery outcomes. They admitted that 'it is not possible for us to separate incremental improvements in outcome due to public reporting from those obtained simply by collection of the data'.³⁵

Hibbard *et al.*¹⁷⁹ Hibbard *et al.*¹¹⁸

Hibbard *et al.*^{118,179} conducted an experiment to evaluate whether or not public reporting was necessary to stimulate QI activities in hospitals. The study assessed providers' responses to the Quality Counts report in Wisconsin. The report compared performance in 24 hospitals in south central Wisconsin. Two summary indices of adverse events occurring within the broad categories of surgery and non-surgery were included, along with indices summarising three clinical areas: hip/knee surgery, cardiac care and obstetric care. Hospitals were rated as better than expected (fewer deaths/complications), as expected or worse than expected. The data were derived from the Wisconsin Bureau of Health Information inpatient public use data sets.

The authors used a quasi-experimental design to assess the relative impact of private and public reporting. The public report group comprised the 24 hospitals that were the subject of and receivers of the Quality Counts report, which was publicly disseminated. The other 98 general hospitals in Wisconsin were randomly assigned either to receive a private report on their own performance or to a control group which received no report. There were no baseline differences between the hospitals in terms of size or pre-report levels of performance. The initial study assessed providers' views of the reports and their self-reported responses to these reports.¹⁷⁹ The follow-up study¹¹⁸ assessed hospitals' performance in the 2 years following the release of the Quality Counts report.

Six to nine months after the release of the report,¹⁷⁹ the public report hospitals self-reported that they were more likely to have engaged in improvement activities in areas that were publicly reported, such as obstetrics (an average of 3.4 out of 7), compared with the private (an average of 2.5 out of 7) or no report hospitals (an average of 2 out of 7), although the differences were statistically significant only between the public report group and the private and no report groups combined. When only the low-scoring hospitals were considered, low-scoring public-report hospitals showed the highest level of QI activities, while the private-report hospitals showed an intermediate level and the no report hospitals showed the lowest level. Two years later,¹¹⁸ 'about a third' (an exact figure is not reported) of hospitals in the public report group significantly improved their performance, while 5% declined. Twenty-five percent of the private report hospitals showed a significant improvement, and 14% declined. When the authors examined only the hospitals that had obstetric performance scores that were worse than expected at baseline (low performers), the differences were more dramatic but showed the same overall pattern. The authors argue that these findings demonstrate that 'making performance information public stimulates quality improvement in the areas where performance is reported to be low'¹⁷⁹ and that there is an 'added value' to making performance public.

In terms of our theories under test, this is consistent with the theory that private reporting stimulates QI activities and provides some indirect support for the idea that providers respond because they have an intrinsic desire to improve. However, it also suggests that public reporting stimulates long-term improvements in quality over and above private reporting, suggesting the public reporting places additional pressure on providers to improve. It also supports the theory that low-performing hospitals that are exposed to public reporting experience greater pressure to improve than average or high-performing hospitals.

Theory 2 summary

We tested theory 2a, that providers respond to privately fed back performance data. van der Veer's¹⁷⁶ systematic review and Taylor *et al.*'s¹⁷⁷ mixed-methods study provide some support for this theory, but also highlight the heterogeneity of the impact of private feedback. Taylor *et al.*'s¹⁷⁷ study suggests that the mechanisms through which private feedback works are to raise awareness of performance in relation to

peers and to motivate improvement through peer competition, in line with audit and feedback and benchmarking theories. However, these studies also suggest that the impact of private feedback is highly contingent on a range of contextual factors, such as the quality of data, the perceived credibility of the data and providers' trust in the data, the availability of and presentation of locally benchmarked data and the degree to which these data offer indications of what needs to be changed. We explore the influence of these contextual conditions in *Chapter 5*. Here, we focus on exploring whether or not public reporting exerts greater pressure than private feedback on providers to respond, theory 2b.

In relation to theory 2b, that public reporting places increased pressure on providers to respond, *Guru et al.*'s³⁹ results suggest that public reporting had no additional impact over and above private reporting on patient outcomes. In terms of the theories under test, this would suggest that public reporting places little additional pressure on providers. In contrast, *Bridgewater et al.*³⁵ found a decrease in overall mortality rates, suggesting that the public disclosure did place additional pressure on providers to respond. However, both studies used a simple 'before-and-after' study design and could not rule out the possibility that improvements in mortality rates were due to changes and improvements in cardiovascular surgery techniques, rather than the introduction of either private or public reporting. As such, we need to be cautious about accepting the claims made by these studies. In both studies, defining a cut-off point to demarcate a move from private to public reporting was difficult, as such transformations occur not overnight, but over a number of months or even years. The providers in both studies had been exposed to private feedback for many years before it was publicly disseminated, and thus may have already responded to and addressed many areas of poor care, thereby attenuating the introduction of public reporting. Furthermore, private feedback may have created a culture where responding to feedback became integrated into practice, which simply continued following the introduction of public reporting. This indicates that previous experience of performance feedback, even if done privately, may have an important influence on provider responses to later, public feedback. *Hibbard et al.*^{118,179} used a more robust study design, in which different groups were subjected to different forms of feedback at the same time, so other more macro changes in the wider policy and institutional environments were similar. They also had a similar pattern of exposure to any form of feedback over time, so there were no maturation effects that may have influenced the impact of each form of reporting. This study suggested that public reporting served to strengthen or accelerate improvements in quality of care provided, and that *public* reporting does place additional pressure on providers to improve care, particularly those who are poor performers. We now turn to understanding the mechanisms through which this may work.

Explaining how public disclosure works

If public disclosure does place additional pressure on providers to respond, we need to explain the mechanism through which this pressure exerts its influence. As discussed earlier, there are several theories that seek to explain why providers might be motivated to take steps to improve the quality of care in response to the public reporting of performance. The 'market share' theory hypothesises that providers are motivated to respond to the public reporting of performance as a result of fear of losing their market share because the public choose higher-quality hospitals over lower-performing ones in response to information on provider performance. The 'professional reputation' theory asserts that providers respond because they are concerned about their reputation which may be damaged as a result of the public disclosure of their performance. Alternatively, providers may respond because they wish to be as good as or better than their peers. We therefore test these theories.

Theory 3: public reporting of poor performance threatens providers' market share and provides a significant provider response

The 'market share' theory assumes that patients use publicly reported information on hospital performance to select higher-quality hospitals, and choose not to go to lower-performing hospitals. In turn, providers improve the quality of care in response to real reductions to their market share or because they fear or anticipate that public reporting will lead to changes in their market share. If patients do choose higher-quality hospitals, it would be expected that hospitals would experience a change to their market

share following public reporting. However, even if providers do not experience changes to their market share, they may still fear the threat of possible changes. Therefore, there are several elements of this theory that can be tested.

Theory 3a: providers experience a change in their market share following public reporting; theory 3b: providers take steps to improve the quality of care because they are worried about potential threats to their market share

To test these theories, we begin by considering studies that examine the impact of public reporting on quality on what health economists term patient's 'revealed preferences' for hospitals, that is, the market share or use of hospitals. These studies use large data sets to examine the outcome of public reporting in terms of its impact on hospital market share. The studies published up until 2012 were summarised in Totten *et al.*'s²⁷ systematic review. We review their findings and then consider a number of studies in this review that may shed light on the mechanisms through which any changes (or lack of) may occur.

Totten *et al.*²⁷

The authors divided their analysis of whether or not public reporting of hospital quality influenced a hospital's market share into studies focusing on cardiac report cards and studies examining non-cardiac report cards. For the cardiac report cards, the authors identified nine studies; in four^{29,32,180,181} of these, there was 'no impact' of public reporting on a hospital's market share. For the other five^{31,38,182-184} studies, there was 'some' impact on market share but 'the effect was small and did not persist over time'. For the non-cardiac studies, one¹⁸⁵ study found a small but statistically significant decrease in discharge rates (used as an indicator for market share) for hospitals with higher-than-expected mortality rates, while two^{186,187} studies reported little or no effect. A further three¹⁸⁸⁻¹⁹⁰ studies found small decreases in market share for lower-rated hospitals or for hospitals that did not produce public reports.

Thus, these studies present a mixed picture of whether or not public reporting of performance leads to changes in a hospital's market share. We now look at some of these studies in greater detail to explore when and how changes in market share did or did not occur. Many of these studies have focused on one of the first and most comprehensively studied report card systems in the US New York State Cardiac Reporting System (NYSCRS). This reporting system has been associated with a decline in risk-adjusted mortality rates for cardiac surgery following the publication of the reports.³⁰ We also pay particular attention to the authors' explanations of why and how changes to market share may have occurred, with a particular focus on why patients may or may not have moved hospital and why providers may or may not have responded to these changes.

Chassin¹⁹¹

Chassin¹⁹¹ compared the market share of high-performing hospitals before and after they were labelled as high performers by the NYSCRS between 1989 and 1995 and the market share of low performers before and after they were labelled as low performers for the same years. To do this, he compared the percentage of all coronary artery bypass graft procedures performed at hospitals classified as outliers for the year before they were labelled as such by the report with the year after. He found either very small or no changes in the market share of groups of hospitals after they were labelled as either high or low performers. He concluded that improvements in New York occurred owing to individual hospitals and cardiac surgery programmes using the data to make specific changes in the way they provided care to coronary artery bypass graft patients.

Chassin cited these findings as evidence that managed care companies (e.g. health maintenance organisations) did not use the data to reward well-performing hospitals or to steer patients towards them. He also argued that patients did not actively search for hospitals with low mortality or avoid those with high mortality. To bolster his argument he presented five case studies to illustrate how different providers made changes to the provision of care in response to being identified as poor performers. Chassin¹⁹¹ highlighted that the providers who took steps to improve the quality of patient care were limited to those who were performing poorly. He attributed the motivation for their efforts to make improvements as the

'opprobrium attached to being named as poorly performing outliers'. In terms of the theories under test, Chassin's argument is that because hospitals did not see a change in their market share following the release of the report cards, it is unlikely that concern about losing their market share was the motivation behind poor performers taking steps to improve care; rather, it was concern about damage to their reputation. However, the author did not directly ask hospital leaders in low-performing hospitals what their motivation for improving patient care was.

Mukamel and Mushlin¹⁸²

Mukamel and Mushlin¹⁸² also examined changes in the market share of providers following the publication of the New York cardiac surgery report cards. They calculated both the absolute difference and the ratio between the market shares of each of 30 hospitals before and after the release of the report cards. Unlike Chassin,¹⁹¹ they restricted their analysis to only patients who were not part of a health maintenance organisation, as patients in health maintenance organisations are often restricted to certain hospitals and thus their choices are constrained. Therefore, their analysis included only patients who, in theory, were free to choose their hospital. Using regression analysis, they examined whether or not there was a relationship between the change in market share for each hospital and their mortality rate. Caution must be exerted in interpreting their findings, as their sample size was small and they did not account for any potentially confounding variables in their analysis. Furthermore, correlation is not evidence of a causal relationship.

Mukamel and Mushlin¹⁸² found an inverse, but non-statistically significant, relationship between the growth rate of a hospital's market share and their reported risk-adjusted mortality rates for each occasion the report was published, with the strongest association being found following the first publication of the reports. They also found a stronger relationship between the growth rate in a hospital's market share in upstate New York than in New York City. They provide two possible explanations for these findings. The first is that patients in upstate New York were more educated and therefore better able to understand the report cards, and also more affluent and therefore more likely to bear the expense of shifting to a higher-quality surgeon. The second is that the New York City market was more competitive and had more information on the relative performance of surgeons, through word of mouth or professional knowledge, to start with. In terms of the theories under test, this study suggests that if report cards do not bring any 'new information' about quality to consumers, consumers will not use them to change providers and, consequently, there will be fewer shifts in market share.

Dranove and Sfekas¹⁸⁴

This study by Dranove and Sfekas¹⁸⁴ tested the hypotheses that when hospital report cards align with prior expectations about quality, there is no change in market share. However, when hospital report cards provide 'news' or new information about quality, there will be movement in market share. Using data from 18 New York State Hospital inpatient records between 1989 (before the report was released) to 1991 (just after the first report was released), the authors developed an econometric model to understand trends in the market shares of these hospitals over time. It is important to bear in mind that quality was measured only in terms of mortality and incorporated only one patient group (cardiac), whose preferences and behaviour may not be generalisable to others. The authors attempted to control for trends in market share by creating a time-lagged variable, although it is unclear how successful this was. It is also unclear if other information excluded from the analyses might have influenced choice and market share, for example news items about health-care litigation cases or department investments (such as new scanners and facilities or expansion activities).

Their findings indicated that when hospital report cards provide information that differs from patients' prior beliefs, patients respond by moving to the highest-quality hospital. They also found that this effect was primarily a result of patients shifting away from hospitals with 'negative news', rather than shifting towards hospitals with 'positive news'. In terms of the theory under test, this suggests that hospitals that previously had a 'good' or even 'average' reputation and are then rated as 'poor performers' by report cards are most likely to have their market share threatened by public reporting.

Theories 3a and 3b summary

The studies here suggest that, overall, there is little change to a hospital's market share following the public disclosure of hospital performance. Although all of the studies examined changes in market share in New York, they came to subtly different conclusions. The studies used different ways of calculating market share and included different populations. Chassin¹⁹¹ analysed market share by groups of hospitals, which may not have been sensitive to changes in the market share of individual hospitals. He also included all patients, including members of health maintenance organisations, whose choice may have been restricted. Mukamel and Mushlin¹⁸² restricted their analysis to patients who, in theory, would be free to move. Although Mukamel and Mushlin did find an inverse relationship between market share and performance, this was not statistically significant. Dranove and Sfekas¹⁸⁴ restricted their analysis to an examination of market share changes immediately after the publication of the first New York report cards, which was also when Mukamel and Mushlin found the most significant impact on market share. Dranove and Sfekas found that it was whether or not performance data provided 'new' information to consumers that influenced their impact on market share.

It is important to highlight a number of caveats when interpreting the findings of these studies. Many of these studies are old, relate to one type of decision/condition (cardiac surgery) and tend to focus on only one or two reporting systems (e.g. New York) and therefore may not reflect patient responses to currently existing public reporting systems. Furthermore, they were conducted in the USA, where the context of patient choice is different from that in the UK. Finally, the studies focused on outcomes to explore whether or not patients actually used the information to inform their choice of hospital.

Nonetheless, in terms of the theory under test, these findings suggest that a hospital's market share is under greatest threat when the market is first informed that hospitals with a previously 'good' reputation are poor performers. As such, this implies that if consumers do move away from a provider, and thus pose a threat to the provider's market share, it is because the report card has damaged the reputation of that provider in the eyes of the consumer. This suggests that the 'reputation' and 'market share' pathways are not necessarily distinct or independent motivations driving professionals to respond to public reports of their performance, as damage to professional reputation can also influence market share. For these reasons, providers may be most concerned about threats to their reputation, rather than about their market share. However, none of these studies has examined providers' perceptions on whether public disclosure of performance represents a threat to their market share or to their reputation. It is studies examining this to which we now turn.

Theory 3c: providers perceive that report cards pose a threat to their market share

We now review studies that have examined providers' perceptions of whether or not performance data have posed a threat to their market share. We start by examining studies based on self-report surveys of clinicians' and hospital managers' views. It is important to bear in mind that self-report surveys are subject to recall and social desirability bias. It is also likely that those who responded to such surveys may have strong positive or negative opinions, which may have influenced the findings. We must therefore exercise some caution in interpreting the claims made by authors on the basis of survey findings. Nonetheless, taken together, the studies provide some useful evidence for testing our theories.

Hibbard *et al.*,¹⁷⁹ Hibbard *et al.*¹¹⁸

The studies conducted by Hibbard *et al.*,^{118,179} discussed earlier, also explored the impact of public reporting on market share and providers' views of whether they were concerned about losing their market share or wished to protect their reputation. The initial study assessed providers' views of the reports¹⁷⁹ and the follow-up study conducted 2 years later¹¹⁸ also assessed changes in market share for hospitals in the public report group by using claims data to compare the proportion of discharges from different hospitals 1 year before the release of the report and 1 year after the release of the report. They also conducted a survey of both community respondents and employees prior to the release of the report to assess their knowledge of hospital quality and resurveyed both groups immediately after the release of the report and 2 years after. It is these findings we consider here.

Six to nine months after the initial release of the report,¹⁷⁹ the performance scores of private and no report hospitals were unrelated to what respondents thought a report would do to their hospital's public image or their market share. Most hospitals considered that it would neither damage nor enhance either their reputation or their market share. However, among the public report hospitals, those with poor scores were more likely to indicate that the report would detract from their public image, whereas those with good scores were more likely to say that the report would enhance it. However, the performance scores of public report hospitals were not related to the anticipated impact of the report on their market share. In terms of the theories under test, this suggests that hospitals whose performance was publicly reported were concerned about their reputation, rather than about their market share.

The follow-up study also found no significant changes in the market share of public report hospitals before or after the report, suggesting that consumers did not change hospitals in response to public reporting. Despite this, consumers exposed to the report were more likely than those not exposed to the report to have an accurate understanding of the relative quality of different hospitals. In terms of our theories under test, this supports the idea that public reporting does increase consumers' knowledge about the relative quality of hospitals and, as such, may affect a provider's reputation. However, consumers may not act on this information and, in turn, this serves to reduce a provider's perceived threat to their market share. Furthermore, the authors also noted that there was little existing market competition between providers, which may also explain why providers felt that only their public image, and not their market share, was threatened. For both reasons, providers did not see the threat of market competition as significant, nor was this the primary motivation for providers to respond to performance data.

Tu and Cameron¹⁹²

This study reports on a survey of hospitals in Ontario 1 year after the publication of the first cardiac mortality report card, the Institute for Clinical Evaluative Sciences Cardiac Atlas, that was made available to the public in 1999. Prior to this date, the data were fed back privately to hospitals. As such, the survey provides a snapshot of the initial reaction of hospitals following the public release of the report card. The report card contained information on 12 acute myocardial infarction performance measures, and covered outcomes (e.g. 30-day risk-adjusted acute myocardial infarction mortality rates, 1-year post-myocardial infarction readmission rates) and process measures (e.g. 90-day post-myocardial infarction beta-blocker rates). A survey was sent to 121 hospitals in Ontario and complete responses were received from 51 doctors (41% response rate); 55% of respondents were chief of cardiology or medicine, while 28% were intensive care unit or coronary care unit directors. The survey contained 24 questions that explored doctors' views on the utility of the different performance measures in the atlas, their limitations, their coverage in the media and the ways in which hospitals had responded to the release of the information. Here, we focus on the findings relevant to testing the theory that providers perceived the public release of performance data to be a threat to their market share or reputation.

Seventy-nine per cent of respondents felt that the Cardiac Atlas had had no impact on the reputation of their hospital, while 15% considered that their hospital's reputation had improved and 6% felt that it had been harmed. Eighty-four per cent of respondents considered that the proportion of patients going to their hospital after Cardiac Atlas publication had remained the same, while 4% felt that it had increased and 12% did not know. None of the respondents felt that the percentage of patients going to their hospital had decreased. Eighty-one per cent reported that none of their patients had discussed the atlas with them and 19% reported that < 10 patients had discussed the atlas with them. Overall, 65% supported the public release of hospital-specific acute myocardial infarction mortality data.

In terms of the theories under test, these findings suggest that doctors in Canada did not perceive the public release of performance data to be a threat to their market share or their reputation. This may be because the survey was conducted only 1 year after the public release of the report card and, as such, the full impact of the report card on market share or reputation had not yet been felt by clinicians. Furthermore, the study findings may also have been affected by selection bias, as only 41% of questionnaires were completed and it is unclear if non-responders had a different perspective from that of those who did respond. Surveys in

general can also be influenced by recall bias, in which respondents' recall of events is partial or influenced by their current situation, beliefs or attitudes.

Guru *et al.*¹⁹³

This study reports on a subsequent survey sent to cardiac surgeons in Ontario, Canada, in 2003, 4 years after the public disclosure of the Cardiac Atlas, described above in the study by Tu and Cameron.¹⁹² The survey was sent to all 55 practising cardiac surgeons in Ontario, and 52 (95%) responded. The survey included both closed Likert scale questions and open questions and explored surgeons' attitudes towards report cards and their beliefs about the impact of report cards.

Eighty per cent of respondents felt that public reporting was important in influencing patients choosing a cardiac surgeon and 84% felt that public reporting was important in influencing the referral patterns of cardiologists. Overall, 51% supported the public release of hospital-specific outcomes and 26% supported the public release of surgeon-specific outcomes. In responding to the open-ended questions, some surgeons expressed positive views about the role of the public disclosure of report cards in enabling public accountability and patient choice: 'Public [have the] right to know and choose improved quality so surgeons cannot hide behind false impressions.'¹⁹³ However some also expressed the opinion that surgeons, rather than the public, should receive feedback so that they can improve care, as the public do not understand report cards. As one respondent commented:

*It is more important for the surgeon and department to be aware and put in place 'checks' to improve quality. Lay people see only a percentable mortality and assume that is either or good or bad without further understanding.*¹⁹³

They also expressed concern that damage to an organisation's reputation can outlive problems with poor care; as one respondent commented:

*Has potential detrimental effects that can be long-lasting even after issues have been corrected.*¹⁹³

It is not clear from the paper how widespread these concerns were among participants; in their discussion, the authors indicated that the responses to open-ended questions 'mainly focused on the negative implications of public reporting'. The respondents and questions in this survey are different from those included in Tu and Cameron's study¹⁹² and differences in question wording can have a significant influence on responses. Despite these caveats, in terms of the theories under test, these findings suggest that, in Ontario, concerns about the impact of public reporting on both market share and professional reputation had increased over time following a switch from private to public reporting of performance. This may be a function of the history of the programme in this setting: clinicians in Ontario had previously been accustomed to the private reporting of performance.

Theory 3c summary

These studies suggest that providers believe that the public reporting of hospital quality poses a threat to their reputation. Providers who have previously been exposed to private reports are concerned about threats to both their reputation and their market share. We need to explore whether providers perceive there to be connection between reputational damage and their market share, or whether these two outcomes are not connected.

Theory 4: providers perceive that report cards damage their professional or their hospital's reputation

The analysis of the studies reviewed in the previous two sections suggested an important revision to our initial theories about the mechanisms through which the public reporting of performance might work. They suggested that the 'market share' and 'professional reputation' theories were not necessarily

independent of each other and that providers were concerned that their market share could be adversely affected if their reputation was damaged. They also suggested that there may be important contextual factors that may limit the threat that public reporting has on an organisation's market share. Here, we review a number of case studies to test the theory that providers are concerned about damage to their reputation because this may in turn affect their market share. We test a number of theories focusing on providers' perceptions that public reporting damages their reputation.

Theory 4a: providers perceive that report cards affect market share by damaging a hospital's reputation

Hildon et al.¹⁹⁴

Hildon et al.¹⁹⁴ conducted a qualitative study of providers' (and patients', although we do not consider these here) views of PROMs data collected by the national PROMs programme in England. Seven focus groups were conducted with a total of 107 clinicians, including consultant surgeons, junior doctors, nurses and allied health professionals. Some caution must be exercised in interpreting the findings; focus groups can be subject to 'chatty bias', whereby those with strong opinions or those higher up in the professional hierarchy may be more likely to make their views known and silence the views of those with different opinions. Furthermore, the focus of the study was to elicit providers' views of different presentation formats of PROMs data; however, the facilitator allowed participants to express broader opinions about the role of PROMs data in QI. As such, the study provides a useful insight into how clinicians perceived the public reporting of PROMs might impact on their professional reputation.

Clinicians were anxious that such data might be reported in the media and misrepresented by both the press and politicians and, thus, might be misunderstood by patients. Clinicians observed:

*You have to remember a politician's going to play with these . . . the Sunday Times and the politicians are going to mess about with them.*¹⁹⁴

*You've got to be very careful because any data that you give out will be interpreted by people who don't necessarily understand what it all means.*¹⁹⁴

Clinicians also expressed grave concerns about the impact on their professional reputation of working in a trust that had been labelled a poor performer by PROMs data. Clinicians commented:

If I believe that I'm providing a five star service . . . then I'm stuck with the label because I'm in this Trust that has been labelled as a three star operation . . . You may be a very good surgeon in that Trust . . . but you're just performing low down because your colleagues brought you down. How do you protect yourself against something like that?

*Reproduced from Hildon et al.¹⁹⁴ with permission from John Wiley and Sons.
© 2012 Blackwell Publishing Ltd*

The damage to their reputations was also perceived to have an impact on their private practice, as one clinician expressed:

*If I feel I'm performing above two stars . . . then I would feel that I'm being hard done by . . . It's a serious issue . . . if people's livelihoods depend on this. Their private practice depends on this.*¹⁹⁴

In terms of the theory under test, this study suggested that clinicians perceived that the data may be misrepresented by the press and politicians, and misunderstood by patients. They were concerned about the potential damage to their professional reputation if they were working in a hospital labelled a 'poor performer' by PROMs data. In turn, they perceived that this may impact on their private practice rather than on NHS patients choosing or not choosing to attend their hospital.

Mannion *et al.*¹⁸

This study examined the impact of the NHS hospital 'star ratings' on acute hospital trusts in England. Recall from *Chapter 3* that the star ratings were a single summary score of hospital performance based on a hospital's achievement on a range of indicators and were made publicly available. Hospitals achieving three stars were judged to have the highest level of performance, two-star hospitals were performing well overall but not consistently in every area, one-star hospitals were a cause for concern and zero-star hospitals had the lowest level of performance against government targets. The authors used a multiple case study design with purposeful sampling of high-performing ($n = 2$) and low-performing ($n = 4$) trusts based on 2000–1 performance data. They undertook documentary analysis (CHI reports and internal governance reports) and semistructured interviews with between 8 and 12 key managers and senior clinicians in each site. As such, their interview findings reflect the views of senior rather than frontline staff. Here we focus on their findings in relation to the perceived impact of the star ratings on a hospital's reputation.

One low-performing trust reported that they had been subjected to a hostile local media campaign that had resulted in adverse public reaction and an erosion of public confidence in the trust, as a quotation from a participant from this trust reveals:

[The public perception was] 'you go to [trust B] and you die!' We had people on the wards demanding the self-discharge forms and getting crushed in the rush to leave! It was just awful. Nurses demanding changing rooms because they didn't want to go outside the trust [in uniform] because they were being accosted in the streets . . . And in the shops, people were saying 'God, you don't work for that place do you? How many have you killed today?'

Mannion R, Davies H, Marshall M, Journal of Health Services Research and Policy (vol. 10, issue 1), pp. 18–24, copyright © 2005 by Sage Publications. Reprinted by permission of Sage Publications, Ltd¹⁸

In terms of the theories under test, this quotation suggests that participants in this trust perceived that damage to their reputation as a result of the star ratings had led to patients wanting to avoid the organisation. The authors found that low-performing trusts also perceived that the poor star ratings had impacted on the hospital's reputation and had affected both the reputation of staff within the trust and the hospital's ability to recruit staff:

I think it [the star rating] gives a very negative view to staff, because if they are working for a one star organisation then it affects the sort of staff who want to come and work for you, but it also makes people who are currently employed here feel that they are working for a third class organisation.

Mannion R, Davies H, Marshall M, Journal of Health Services Research and Policy (vol. 10, issue 1), pp. 18–24, copyright © 2005 by Sage Publications. Reprinted by permission of Sage Publications, Ltd¹⁸

This suggests that the star ratings were perceived as affecting the hospital's reputation not only in the eyes of the public, but also in the eyes of staff working within that organisation and those of peers in other hospitals. This highlights that a hospital's reputation depends not only on the opinions of patients, but also on those of other clinicians (i.e. their peers).

Mehrotra *et al.*¹⁹⁵

Mehrotra *et al.*¹⁹⁵ conducted interviews with 17 employers and 27 hospital managers to explore their views of and responses to employer-initiated report cards within 11 regions of the USA. The authors attempted to include hospital representatives who were supportive of and those who were opposed to report cards. Hospital managers were either chief executives or QI directors.

The authors explored the theory that the public release of report cards would capture the attention of the media and consumers and that providers would respond by improving the quality of care. They found that, in most instances, the initial public release of performance data had led to resentment, as hospitals were

unhappy that data 'they perceived to be inaccurate' were shared with consumers. Their respondents felt that media attention had been patchy and tended to focus on 'controversies' such as a hospital with a previously good reputation performing poorly. Some participants felt that they were being 'beaten over the head' by the press and were concerned that this would scare consumers and lead to them 'stampeding the doors wanting change or getting politicians involved'.¹⁹⁵ They found that, in general, participants in communities that already had public report cards were less fearful of public disclosure than those in which feedback was conducted privately. Participants who had experienced public reporting felt that, after the initial wave of media attention following the first release of the report, coverage decreased over time and consumer interest in the report was low. However, some participants felt that this low level of consumer interest was still sufficient to prompt change as it gave 'hospitals time to fix their problems without horrible penalties . . . But if they ignore it for five years, all of sudden you're looking at a more significant market share shift'.¹⁹⁵

The authors also found that hospitals preferred reimbursement (i.e. financial rewards) for high quality rather than employers directing their employees towards 'better' hospitals and thus increasing their market share. Chief executives felt that, although employers could threaten to do this, they could not ultimately control where patients chose to go. Furthermore, even if they could, many hospitals were full, and so patients would not be able to attend the high-performing hospital in any case.

In terms of the theories under test, these findings suggest that those who experienced private feedback were more fearful of the impact of public reporting on their market share. This may explain why Guru *et al.*¹⁹³ found an increase in concerns about public reporting when it was introduced following a period of private reporting. They also indicate that media coverage which focused on 'controversies', such as when a well-regarded hospital had a poor public report card, was the means through which public reporting had the potential to damage a hospital's reputation. Participants perceived that this in turn would lead to consumers leaving their hospital. This reinforces the findings of Dranove and Skekas,¹⁸⁴ that it is when report cards bring 'new' information to consumers that they impact on market share. However, the findings also suggested that such an effect may be short-lived and that, in general, consumers were not interested in public reports. Furthermore, they also suggested that the 'threat' that providers could lose their market share was an empty one if 'better' hospitals were full.

Greener and Mannion¹⁹⁶

This paper presents an ethnographic case study of one hospital trust in the north of England that was part of a larger project to examine cultural change in the NHS. Although the paper is primarily focused on providers' views of the patient choice agenda in the NHS, findings also touch on providers' perceptions of the impact of performance data. Data were collected over a period of 2 years (2006–8) and included field notes from observations of meetings and 'hospital life more generally' and 60 interviews with staff from different levels of the organisation. Their findings shed light on providers' responses to attempts by the government to increase competition between providers and on their perceptions of how GPs and patients have responded to increased information about hospital performance.

The authors found that hospital managers perceived GPs, rather than patients, to be their real 'customers'. Hospital managers perceived that GP referral patterns were 'long established' and largely governed by 'history and tradition', rather than being influenced by data on mortality statistics. For example, one participant observed:

So people still want to go to their local hospital, and . . . you know, there weren't, despite increasing availability of statistics on mortality and all those things, people still went, it wasn't a discerning factor.
Senior service manager¹⁹⁶

Managers also perceived that patients were loyal to local services and wanted to go there rather than anywhere else. A senior medical manager observed that there was 'a huge amount of loyalty in the local

population for local services¹⁹⁶ and that this persisted even in the face of competition from independent sector treatment centres that provided similar services. One manager explained:

The independent sector treatment centre programme . . . cherry picked, you know: we'll do simple hips and things and we'll do simple cataracts and cream off a bit of money. Very mixed experience, people still want to go to their local NHS for it.

Senior service manager¹⁹⁶

This quotation suggests that patients remained loyal to their local NHS hospital and that independent sector treatment centres accepted only less complex patients, thus limiting the perceived competition they posed. Furthermore, managers commented that the demand for services was so high that they were more concerned with limiting the number of patients admitted than with losing patients. As a board member of the trust explained:

To be completely honest, we're so flush with demand that more of our conversations are about how we can . . . constructively decline . . . patients than anything else.

Board member¹⁹⁶

In terms of the theory under test, a notable contrast between this case study and both Mehrotra *et al.*¹⁹⁵ and Mannion *et al.*¹⁸ was the lack of media coverage about public reports of mortality statistics. Under these circumstances, the availability of information on hospital performance was perceived as having little impact on the choices of either patients or GPs. Instead, these choices were seen as being driven more by habit and tradition than by external data. Furthermore, this study suggests that other contextual factors, such as a high demand for services and independent sector treatment centres (the 'competition') accepting only less complex patients, had also limited the extent to which hospital managers may have felt that performance data posed a threat to their market share. Indeed, it was dealing with too many patients, rather than having too few, that was their main problem.

Theory 4a summary

These studies lend some support to the theory that providers are concerned about the damage to their reputation as a result of the public reporting of their performance, as they perceive that this may, in turn, affect their market share. Both Mehrotra *et al.*¹⁹⁵ and Mannion *et al.*¹⁸ found that their participants made a direct connection between the effects that public reports of their performance had on their reputation and their concerns that patients would want or wanted to leave their hospital. Hildon *et al.*¹⁹⁴ found that clinicians feared that damage to their reputation would affect their private practice. The authors also suggested that reputational damage rests in the opinions not just of consumers but also of clinical peers. However, these studies also suggest a further revision to our theory: that media portrayal of the hospital was the means through which damage to the hospital's reputation came about. More specifically, Mehrotra *et al.*¹⁹⁵ suggested that the media coverage focused on 'controversies'; that is, instances in which a previously well-regarded hospital received a poor report card were more likely to be reported. The studies also suggested that, without this, consumers and those who may refer them (such as GPs in England) were not influenced by report cards, as Greener and Mannion's study¹⁹⁶ discovered. Furthermore, the studies also indicate that threats to a provider's market share and the fear that consumers may abandon low-performing hospitals, in favour of higher-performing hospitals was reduced when GPs and patients were loyal to local hospitals and when demand was high and the capacity of better-performing hospitals to accept additional patients was limited. We now explore, in more detail, the role of the media in influencing a hospital's reputation and market share.

Theory 4b: media reports of hospital performance damage a hospital's reputation and affect their market share

Pawson,²¹ in his review of public disclosure programmes, drew attention to the role of the media in publicising the information contained in public reports and drawing the public's attention to them, thus triggering public sanctions, for example a shift away from poorly performing hospitals. The HCFA mortality

report cards provide a useful case study to understand how the media reporting may influence a provider's reputation and its market share. The HCFA reports were one of the first reporting systems in the USA. Hospitals were categorised as having a lower than expected, as expected or higher than expected mortality rate for nine patient groups. It is important to note here that the reports were not originally intended for use by the public, but press and consumer groups forced their public release under a Freedom of Information Act in 1987, and the reports were then released to the public until they were stopped in 1992 owing to concerns about the validity of the risk-adjustment methodology.¹⁹⁷ This little tale speaks a lesson in itself, which we will consider further in subsequent sections of this report. Here, we review a number of studies that have examined the impact of these report cards on market share and how media reporting of them may (or may not) have contributed to this.

Vladeck *et al.*¹⁸⁶

Vladeck *et al.*¹⁸⁶ examined the occupancy rates, as an indicator of market share, for all New York City hospitals labelled as having mortality rates above, below or at the predicted levels according to the HCFA data. They compared occupancy rates for the five calendar quarters preceding the release of the reports with occupancy rates for the three calendar periods following the report's release. They found no statistically significant differences in occupancy rates between the two periods. They also found that hospitals with lower-than-expected mortality rates did not experience an increase in market share, and that hospitals with higher-than-expected mortality rates did not experience a decrease in market share; in fact, the authors found an opposite, albeit not statistically significant, trend. The authors concluded that the public release of mortality data does not discourage consumers from utilising poorly performing hospitals.

In their discussion, the authors speculated two possible reasons for their findings. The first was that the version of the HCFA 'death list' they studied was 'seriously methodologically flawed'¹⁸⁶ and, as such, consumers might have dismissed its usefulness as an indicator of the quality of hospital care. However, one might question if consumers would have had the necessary knowledge to discern that performance data were flawed. The second is that patients' choice of hospital more influenced was by 'preferences for and by doctors, tradition, convenience and word of mouth . . . than with objective information about hospitals'.¹⁸⁶ In other words, consumers rely more on a hospital's reputation in the eyes of friends, family and referring clinicians than on 'objective' information about hospital quality.

Menemeyer *et al.*¹⁸⁵

Menemeyer *et al.*¹⁸⁵ examined discharges (as an indicator of market share) from US community hospitals with a standardised HCFA mortality rate of more than one standard deviation from the mean in any year and 50% of the hospitals that were never outliers between 1984 (2 years before the first HCFA report was released) and 1992. They also collected data about newspaper reports of hospital quality from a full-text online retrieval service, and categorised these according to whether the stories related to the hospital being a high or low outlier on the HCFA data and also to stories unrelated to the HCFA data, such as an 'unfavourable story' or an 'untoward death'. They used a number of econometric models to examine the relationship between hospital discharges and the release of the HCFA reports and the media stories. They found that hospitals with higher-than-expected mortality rates did experience a small but statistically significant reduction in discharges following the release of the HCFA data. Their study also assessed whether or not subsequent press reports of the HCFA data affected the number of discharges but did not find any statistically significant impact. However, they did find evidence of a large and significant impact of press reports of 'untoward deaths' on market share (9% reduction in hospital use).

Menemeyer *et al.*¹⁸⁵ concluded that the release of the HCFA data had only a small but statistically significant effect on market share and found no support for the theory that media reports of the data influenced market share, although they included the caveat that 'limited data availability keeps us from making too much of this finding'. They argued that the public 'essentially ignored a sophisticated attempt to measure quality' but instead took notice of 'very simple, unsophisticated . . . stories about untoward deaths'.¹⁸⁵ Pawson²¹ argued that this study shows that the information contained in the report cards regarding hospital performance did not reach the public's consciousness, neither via the public release of the information nor via the media's reports of

the data, but sensational stories about 'tales of the unexpected' did. Menemeyer *et al.*¹⁸⁵ noted that perhaps one reason the media reports of the HCFA data had little impact on consumers was that newspaper reports of the HCFA release often included interviews with a hospital leader who discounted the results owing to the faulty analysis by the HCFA. This suggests that the content of the media reports about public report cards may determine their impact on a hospital's reputation and, consequently, on their market share. It is this issue that we turn to next.

Rudd and Glanz¹⁹⁸

Rudd and Glanz¹⁹⁸ analysed newspaper coverage of the HCFA hospital mortality data to examine the type, amount and themes of newspaper coverage of the HCFA mortality data and whether or not this was dependent on the ratings of a hospital's quality. They took a convenience sample of 68 newspaper articles (published in newspapers serving 47 small, medium and large urban areas from 28 regions across the USA), 60 of which were published the day after the release of the reports (18 December 1987). The authors analysed where the reports appeared in the newspaper, the number of lines the reports received, whether the headline was positive, negative or neutral and the number of hospitals mentioned in the report that had higher-than-expected, lower-than-expected or as-expected mortality rates. They also examined the explanations offered by representatives of hospitals with higher than average mortality rates.

Their analyses indicated that the HCFA reports received high newsplay, with almost half of the stories appearing on the front page. They found twice as many negative headlines as positive headlines. Forty-one percent of the articles had negative headlines that emphasised higher-than-expected mortality ratings, 42.6% had neutral headlines and 16% had positive headlines. Articles with negative headlines were no more likely than articles with positive headlines to mention a high number of hospitals with higher-than-expected mortality ratings. In other words, the headline did not necessarily directly reflect the information contained in the articles. They also found that articles over-represented hospitals with higher-than-expected or lower-than-expected mortality rates and under-represented hospitals with as-expected mortality rates. The most common sources of quotations found in articles were from hospital and medical representatives. The most frequently cited explanations for higher-than-expected mortality rates were to blame some aspect of the HCFA methodology (69% of explanations) or to blame some aspect of the case mix of their patients (59% referred to the patients' illness levels and 35% to their social characteristics).

The authors concluded that press reports of the HCFA data 'no doubt raised consumer awareness of the concept of healthcare quality' but also 'planted doubt about the so-called "hard" data on quality of care'.¹⁹⁸ They also argued that newspaper reports did not provide 'concrete guidance on how to obtain useful and valid data on health care quality'.¹⁹⁸ In terms of the theory under test, these findings suggest that the public release of performance data may raise awareness of health-care quality but does not necessarily provide consumers with valid and reliable information to inform their choice of hospital, as the data tend to focus disproportionately on those who are either below or above average. Furthermore, the findings also suggest that these reports might send the message to consumers that these data cannot be trusted.

Berwick and Wald¹⁹⁷

Berwick and Wald¹⁹⁷ conducted a survey of hospital leaders following the release of the HCFA data. From each state, they randomly selected two hospitals with higher than expected mortality rates, two with expected rates and two with lower than expected mortality rates, giving a total of 250 hospital in their sample. Each hospital was sent a 12-item survey exploring hospital leaders' opinions of the accuracy and value of the HCFA data, whether or not they had used the data and the degree and extent of any problems. A total of 195 hospitals responded, giving a response rate of 78%.

Seventy per cent of respondents rated the usefulness of the HCFA data as poor, 54% rated the data's accuracy as poor and 85% rated the data's usefulness to consumers as poor. Hospitals in the high-mortality group were more likely to report that the public release of HCFA data had 'caused problems in their dealings with professionals, purchasers, patients or others'.¹⁹⁷ In response to an open question, the most common problem reported was 'misrepresentation of the data by press and public'.¹⁹⁷ In terms of the theory under test,

these findings suggest that, in line with the quotations from hospital representatives contained in press reports found by Rudd and Glanz,¹⁹⁸ hospital leaders perceived the data to be inaccurate and not useful for their own QI efforts or for patients choosing hospitals. They also indicate that leaders in hospitals that had a higher-than-average mortality rates experienced problems in their dealings with a range of stakeholders as a result of the perceived misrepresentation of the data in the press.

Theory 4b summary

This collection of studies about the HCFA data suggest that the report cards themselves did not lead to patients changing hospitals in preference for 'higher-performing' hospitals, and that patients took little notice of press releases about these reports. Hospital leaders themselves distrusted the data, and this message also found its way into press coverage about the reports. This may have been one reason why the reports themselves or press reports about them appeared to have little impact on patients' choice of hospital. Another is that patients base their decisions on advice from referring doctors, friends and family, and on their own previous experiences. Patients' choices of hospitals were influenced by stories about untoward deaths, or 'tales of the unexpected'. These stories may have better captured the public's attention because they were easier to understand than stories about report cards or because they were simply more interesting to patients. However, it seems that press reports did not go unnoticed by hospital leaders, especially those in hospitals classified by the HCFA data as having higher-than-average mortality rates. They were concerned not only about the validity of the HCFA data but also about its misrepresentation by the press and public. Press reports about the HCFA data disproportionately mentioned hospitals that performed above or below average, so these concerns are not without foundation. In terms of our theories under test, this suggests that hospital leaders may be more concerned about and influenced by the press coverage of reports cards than patients are.

Theory 5: providers respond through comparing themselves with their peers

We now review studies that consider whether or not providers are motivated to implement QI initiatives through a process of comparing themselves with their peers and consequently being alerted to difference between their own performance and that of peers. The motivation to outperform or be as good as peers is the central mechanism motivating organisations to take steps to improve performance within theories of competitive benchmarking.⁹¹ In contrast, theories of collaborative benchmarking focus on the role of learning from the best practices of others as the mechanism through which organisations improve their performance. We examine studies to test the theory that competitive or collaborative benchmarking is a mechanism that underpins efforts to protect professional reputation and public image.

Hildon et al.¹⁹⁴

Hildon et al.¹⁹⁴ conducted a qualitative study of providers' views (and patients' views, although we do not consider these here) of PROMs data. Seven focus groups were conducted with a total of 107 clinicians, including consultant surgeons, junior doctors, nurses and allied health professionals. Some caution must be exercised in interpreting the findings; focus groups can be subject to 'chatty bias', whereby those with strong opinions or those higher up in the professional hierarchy may be more likely to make their views known and silence the views of those with different opinions. Furthermore, the focus of the study was to elicit providers' views of different presentation formats of PROMs data; however, the facilitator allowed and encouraged participants to express broader opinions about the role of PROMs data in QI. As such, the study provides a useful insight into how clinicians perceived the public reporting of PROMs may impact on their professional reputation.

Clinicians recognised the potential of the national PROMs programme to stimulate providers to take steps to improve the quality of patient care. The mechanism through which this was perceived to happen

was a process of benchmarking, whereby clinicians would learn from the practices of better-performing providers. As one clinician commented:

*I'd want to know who's got a service that's better than mine. And then I'd go and visit them and find out what's their secret.*¹⁹⁴

Another clinician commented that they wanted to know how they compared with other colleagues, suggesting that professional rivalry was a motivating factor to improve their performance:

*I'd like to know how high I do and that's what every surgeon would like to know. How do I compare to my colleagues here, next door and nationally?*¹⁹⁴

This study suggests that clinicians find PROMs feedback useful because it enables them to compare their own performance with those of their peers, in line with theories of competitive benchmarking.

Greer¹⁹⁹

All quotations from this study are reproduced from Greer¹⁹⁹ with permission from the Commonwealth Fund.

Greer's¹⁹⁹ study describes why the Wisconsin Collaborative for Healthcare Quality (WCHQ) was set up and how the indicators were developed. This was a clinician-led initiative that was set up to develop their own public reporting programme. The report is based on 31 interviews with those involved in the WCHQ, including board members and medical practice executives. They included eight chief executive officers (CEOs), 10 chief medical officers and five executives responsible for the quality of care in their organisation. As such, the study does not capture the perspectives of staff on the ground or their responses to this initiative. Nonetheless, the article discusses some of the mechanisms through which change was perceived to occur and the contextual factors influencing its success. Greer's interviewees noted:

I think one of the constructs that the Collaborative is built on is that physicians want to do a good job . . . By providing information about how physicians perform you can influence physicians behaviour . . . They are driven to change things when their performance does not look good.

p. 14¹⁹⁹

WCHQ [provides] the actual benchmarking data for looking at where you are at, how to improve, and building on those connections with other organisations that are similar to you; where you can say 'our numbers are not good here, how did yours get better? What can we learn from how you are doing it?'

p. 17¹⁹⁹

Greer's report of how the WCHQ was set up is ostensibly a 'good news story', as the report is overwhelmingly positive about the Collaborative and is based on data from the enthusiasts who led it. The author recognises that, as the collaborative was a 'pioneering effort', it benefited from the Hawthorne effect. However, it does shed some light on the mechanisms through which a voluntary (as opposed to mandatory) public reporting system may work. In terms of the theory under test, these quotations illustrate the idea that clinicians were intrinsically motivated to do a good job, and receiving feedback which indicated their performance was poor prompted them to take steps to improve. The second quotation supports the collaborative benchmarking theory that clinicians can identify what needs to be improved and how by sharing and learning from the best practices of other providers.

Davies⁹⁸

Davies⁹⁸ conducted a multiple case study of six hospitals in California, USA, to examine providers' responses to publicly reported data and to internally produced and confidentially fed back data, with a specific focus on cardiology. The hospitals were specifically selected to be 'high performers', as the authors assumed that

they would be more likely to find examples of QI here (with the assumption that the hospitals had become 'high performers' as a result of QI activities). The data consisted of 35 interviews with 31 participants, including chief executives, senior clinicians with management responsibilities, senior quality managers, the chief of cardiology, a senior nurse manager and two or three frontline staff in the cardiology department.

Their findings suggest that a key motivator of change was 'credible comparative data of quality problems and detailed exploration of clinical processes coupled with professional and institutional pride'. Comments from participants in their study suggest that clinicians were motivated to take steps to improve the quality of care in order to be as good as or better than their peers:

So I do see physicians taking it very seriously, they do want data to reflect favourably on them, there's a tremendous pride in their work.

*Physician*⁹⁸

If you've got the best outcomes [and] the least complications you have a higher standing with your peers. And if you know you've got a problem and you address it, that improves your standing . . . They [physicians] are also very competitive. They want to do the right thing, and they want to do it as well or better than everybody else.

*Chief of staff. Reproduced from BMJ Quality and Safety, Davies HTO, vol. 10, pp. 104–10, © 2001 with permission from BMJ Publishing Group Ltd*⁹⁸

Physicians are self correcting, they're very competitive, they always want to be the best. If you show them data and they're not as good as their partner, they tend to try and figure out themselves what's going on.

*Quality improvement manager*⁹⁸

In terms of the theory under test, these findings suggest that clinicians are seen as intrinsically competitive and that feeding back data that suggest that they are not as good as their peers is what motivates them to improve their own practice. This appears to be in line with theories of competitive, rather than collaborative, benchmarking. The findings also suggest that a clinician's reputation rests on the views of their peers, and that this is an important determinant of who referring clinicians decide to send patients to. As one chief of staff in a hospital that was part of health maintenance organisation noted, 'It's the opinions of peers that matter more than anything else about quality. Who do people go to for consults?'.⁹⁸ Thus, being better than or as good as one's peers improved one's reputation, as did taking steps to improve one's own practice. This, in turn, meant that peers were more likely to refer their patients to you.

Theory 5 summary

These studies provide some support for the theory that providers are motivated to take steps to improve the quality of patient care in order to be as good as or better than their peers. They also suggest that doing so serves to enhance or maintain a provider's reputation which, in turn, means that peers are more likely to refer patients to them. Thus, this suggests that the 'competitive benchmarking' and 'reputation' pathways are interconnected rather than mutually exclusive.

Theory 6: consumers choose hospitals on the basis of information about the quality of services

The assumptions underlying the public reporting of hospital performance data are that patients use quality information to choose a hospital and will change hospitals in response to learning about a poorly performing hospital. However, evidence reviewed in theory 4, regarding professional reputation theories, suggests that providers perceive that patients are loyal to their local hospital unless the media report the hospital's poor performance. We now test a number of subtheories concerning how patients choose hospitals.

Theory 6a: patients are willing to move hospitals in response to poor performance; theory 6b: patients choose hospitals on the basis of reports about service quality; theory 6c: patients respond to poor hospital performance when it is reported in the media

To test these theories, we now turn to considering studies that have examined whether or not patients are willing to change hospital and what factors influence patients' choice of hospital. These studies have used a number of different methodologies to explore this question. Some have used surveys or interviews to examine patients' stated preferences, that is, the factors patients report as influencing their choice of hospital shortly after receiving care or when real-life choices are being made. A strength of this methodology is that studies examine choices actually made by patients in the context of their own lives. However, a limitation is that patient responses can be heavily influenced by question wording and whether patients are asked to recall what factors are important, choose a number of factors from a prompted list or select the one most important factor from a prompted list. As with all survey methods, responses can be subject to recall bias. Other studies have used discrete choice experiments, in which patients are asked to make hypothetical choices about fictional hospitals. The strength of this methodology is that the attributes of the hospitals can be set at different levels by the investigator to explore the relative importance of different characteristics in determining patients' choices. They force respondents to explicitly trade off different attributes of hospitals, rather than allowing respondents to consider each attribute separately. They also enable an exploration of patient choice free from the constraints of the market and reimbursement context which may limit patient choices. However, a limitation of these studies is that patients' hypothetical choices about fictional hospitals may not reflect the choices they make in real life. Therefore, by comparing and contrasting and integrating findings using these different study designs, we can gain a deeper understanding of how patients choose hospitals to understand the relative influence of performance data in these decisions.

Faber et al.²⁰⁰

This systematic review aimed to explore what weight consumers give to quality of care information in choosing providers, to understand how presentation format of quality information influences patient choice, and to examine the influence of quality care information on consumer choice in real-world settings. In our synthesis, we focus on the first and last of their research questions, as these are most relevant to testing the theories set out above. The authors conducted a comprehensive search strategy and restricted inclusion to RCTs, controlled before-and-after trials and interrupted time series. The authors assessed study quality using a validated checklist, and two RCTs were excluded from the review because they were of low quality.

The authors identified five laboratory studies that used discrete choice experiments to examine the weight consumers gave to quality information in choosing a health plan in the USA. This is not the same as choosing a hospital; a health plan is essentially a health insurance policy that varies in terms of coverage of services and costs. Health plans often limit consumers to attendance at particular hospitals and these experiments often included quality information on the hospitals included in the health plan, as well information on costs. The authors concluded from the findings of two studies^{201,202} that 'restrictions of provider choice seemed to overrule the weight given to quality'. One study²⁰¹ found that, in the absence of cost information, information on quality only induced consumers to choose higher-quality rated health plans, and reduced the importance of other features only when there were large quality differences between the plans. Three studies examined the trade-offs consumers made between the cost and quality of health plans. Two studies^{203,204} found that cost data reduced the demand for high-quality health plans, provided that the lower-cost health plans were of high quality. However, another study²⁰² found that within health maintenance organisation-managed care programmes, consumers were less likely to select a low cost plan even when these were also of high quality. Together, these studies suggest that, overall, quality data have little impact on consumer choice of health plans.

The authors also identified four 'real-world' experiments²⁰⁵⁻²⁰⁸ that examined consumers' stated preferences in terms of their actual choice of health providers, following exposure to the Consumer Assessment of Healthcare Providers and Systems (CAHPS) performance data on new Medicaid beneficiaries

or employees who had to select a new health plan. These studies all found no effect of the CAHPS information on consumer knowledge, attitudes or behaviour. However, two studies^{205,207} grouped together all consumers who reported having seen the CAHPS report, irrespective of their study allocation. One study²⁰⁷ found that the subgroup who saw and remembered the CAHPS data selected health plans with higher CAHPS ratings, while the other found that this subgroup perceived the CAHPS information as being more important when choosing a health plan. These findings suggest that whether consumers remember seeing quality data may be an important determinant of whether or not these data inform their choices.

Dixon²⁰⁹

This report provided an overview of findings from the national patient choice survey. A questionnaire was sent to 205,000 patients in England who had been referred by a GP for a first outpatient appointment between 15 and 28 February 2010; 69,000 (34%) responded. Here, we focus on answers to 3 of the 11 questions asked as part of this survey:

Q3: Were you offered a choice of hospital for your first hospital appointment? (Yes/no.)

Q4: Which was the single most important source of information for you when you chose your hospital? (Patients were presented with a list of eight different sources and asked to choose one.)

Q6: Which was the single most important thing for you when you chose your hospital? (Patients were presented with a list of 14 different factors and asked to select one.)

Dixon²⁰⁹

In response to question 3, 49% of patients were offered a choice of hospital. In response to question 4, 43% of patients who were offered a choice of hospital reported that their GP was the single most important source of information when they chose their hospital, followed by 29% indicating that their own experience or that of friends and family was the most important source. Six per cent indicated that a book or leaflet produced by the local PCT, containing performance information, was the most important source, and 4% stated that the NHS Choices website was the most important (*Figure 12*).

In response to question 6, the most important factor affecting choice of hospital for 38% was close proximity of the hospital to their home or work, for 12% was their personal experience of the hospital,

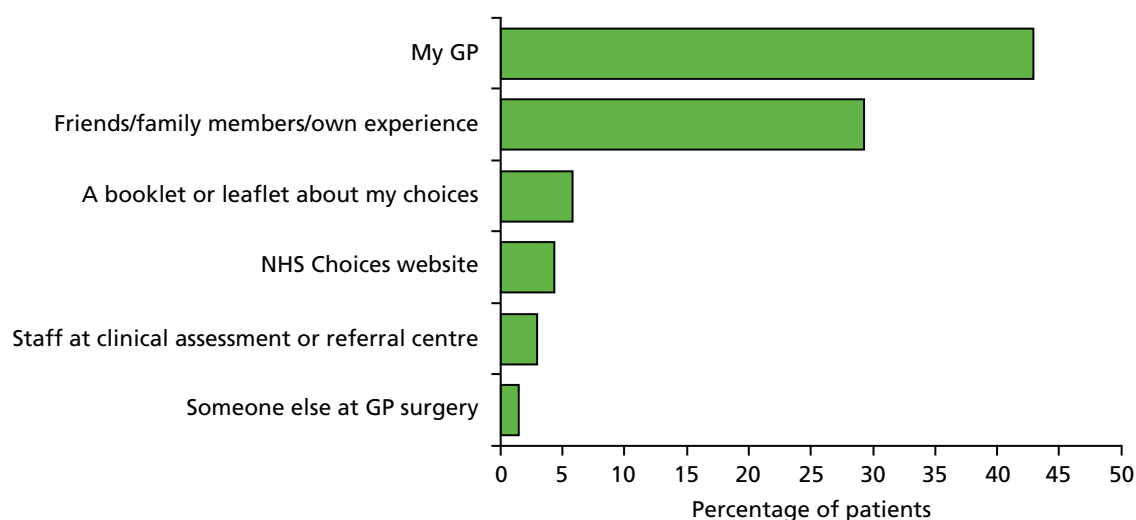


FIGURE 12 Sources of information used by patients offered a choice of hospital. Reproduced from Dixon.²⁰⁹ Contains public sector information licensed under the Open Government Licence v3.0.

for 10% was length of wait for an appointment, for 6% was good previous experience and for 5% was quality of care (Figure 13).

In terms of the theory under test, these findings suggest that just under half of patients referred to secondary care were offered a choice of hospital. The most important source of information for patients offered a choice was their GP, followed by friends and family. Few patients consulted the NHS Choices website that provided comparative information on hospital quality. Patients were more concerned about going to a hospital close to their home or work, and were more influenced by their previous experience of the hospital. This suggests that comparative information on hospital quality did not play a significant role in patients' choice of hospital.

Abraham *et al.*²¹⁰

This paper reported on a survey of patients attending four medical centres in Minneapolis, USA. A total of 699 patients were asked to complete the survey, of whom 467 did (a response rate of 66.8%). Patients were asked to rate how important a number of different factors were in choosing a health-care provider or doctor on a 5-point scale from 'not at all important' to 'very important'. Patients were also asked about their awareness of different formal sources of cost and quality information and, if they were aware of the source, whether or not this had influenced their decision-making.

The factors with the highest ratings of importance were the reputation of the organisation (91% rating this as 'important' or 'very important') and the reputation of the doctor (90% rating this as 'important' or 'very important'). Other factors included appointment availability (72%), referral from a doctor (69%) and a recommendation from family and friends (65%). The least important factors were websites that reported clinical quality data (24.2%) and television, radio and media advertisements (8.97%). The authors reported that 36% of respondents reported being aware of individual health-care providers' websites and, of those who were aware, 5% felt that these had influenced their choice of provider. Of the other websites providing hospital quality and cost information, that with the highest level of awareness was 'Angie's List', with 36% of respondents being aware of this website; of these, 2% felt that it had influenced their choice. The authors noted that this website also contained reviews and recommendations for other services, such as home improvements, in addition to health care, which suggests that respondents might have been aware or accessed this website for other reasons, rather than as a source of health-care quality information. On the basis of these findings, the authors conclude that patients prefer to rely on more informal sources of information, rather than formal sources of performance data, when choosing a hospital. This was a relatively small survey in one location in the USA. However, it provides further support

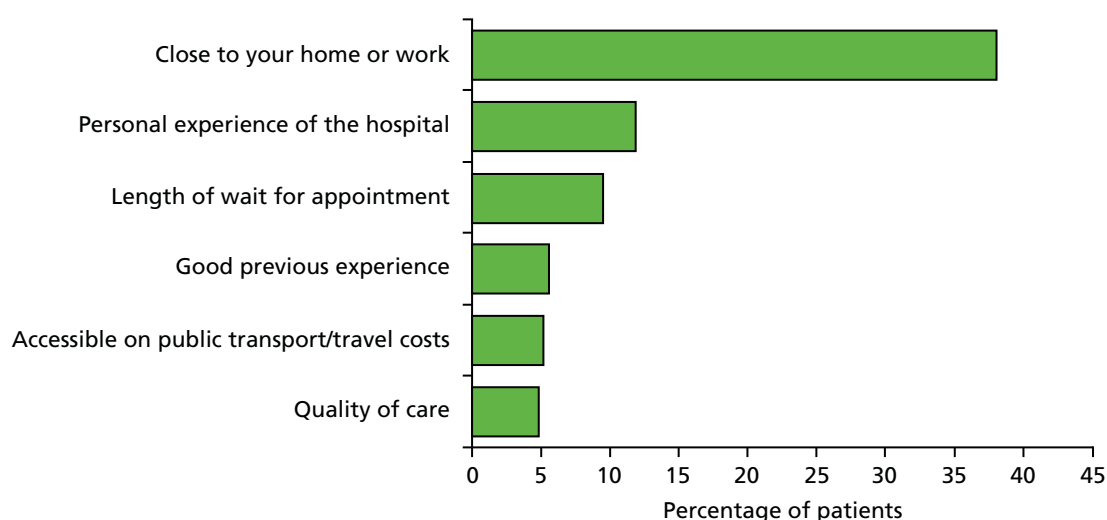


FIGURE 13 Most importance factors for patients offered a choice when choosing a hospital. Reproduced from Dixon.²⁰⁹ Contains public sector information licensed under the Open Government Licence v3.0.

to the findings from Dixon's larger UK survey that patients rarely consult formal performance data when choosing a hospital, but instead base their decisions on a hospital's reputation.

Djis-Elsinga *et al.*²¹¹

This paper reported on a survey of patients who had recently undergone surgery in the Netherlands, to explore the factors that informed their choice of hospital. The sample comprised all 2122 patients who had undergone one of six surgical procedures (aorta reconstruction, cholecystectomy, colon resection, inguinal hernia repair, esophageal resection and thyroid surgery) in three hospitals between 2005 and 2006. A total of 1329 questionnaires were available for analysis, giving a response rate of 62.6%. The survey asked patients to indicate which of 14 information items they had used to inform their choice of hospital; these comprised 11 'general information' items and three items relating to quality of care information.

The most common sources of information patients had used to inform their decision about where to have surgery were the reputation of the hospital (69% indicating that they had used this information) and the friendly atmosphere of the hospital (63%). This was followed by easy access by their own or by public transport (48%), the distance to hospital (44%), good parking facilities (41%), hospital rooms equipped with personal facilities (40%) and the fact that they had already been treated by that hospital/surgeon (40%). The least commonly used sources of information were information items about the quality of care, with 3.2% of respondents indicating that they had considered the percentage of patients with an adverse outcome after surgery, 2.1% who had considered the percentage of patients with little pain and < 1% who had considered the percentage of patients with pressure ulcers. It is not clear from the survey whether patients did not consider quality of care information because they did not have access to the information, because they had access but did not understand it or because patients did not consider this to be important. However, this survey also provides further support to the theory that, when choosing a hospital, patients prefer to rely on 'soft' information, such as the reputation of the hospital or its staff, than on performance data.

Marang-van de Mheen *et al.*²¹²

This study, by the same group of researchers, used a discrete choice experiment to examine the relative importance of quality of care information in patients' choice of hospital. Respondents to the survey, reported by Dijs-Elsinga *et al.*,²¹¹ who had recently undergone one of six surgical procedures, were asked if they were willing to be contacted again to take part in a further study; 665 out of 1329 patients agreed to be contacted, and 559 of these responded to an invitation to participate in the study. Of these, 369 agreed to take part in the study and were sent a questionnaire, and 308 returned a questionnaire.

The questionnaire consisted of a discrete choice experiment, in which patients were presented with 12 different comparisons of two hospitals. Each hospital was characterised by six attributes, which varied in terms of two different levels and were based on responses to a previous survey about which factors patients most commonly considered in choosing a hospital. The attributes included:

- previous experience with that hospital or surgeon (previous experience vs. no previous experience)
- hospital reputation (good vs. less than good)
- hospital atmosphere (personal vs. businesslike)
- waiting time for surgery (2 weeks vs. 6 weeks)
- percentage of patients with a 'textbook' outcome, defined as no readmissions or operations, no adverse outcome reported and a hospital stay that was no longer than average
- a surgery-specific outcome that varied according to the procedure patients had had and focused on the opportunity for minimally invasive surgery (for cholecystectomy and hernia repair)
- surgeon's experience of the procedure (for colon resection and thyroid surgery) and volume of procedures (for aorta reconstruction).

The different attributes and their levels were described to patients in the questionnaire, and they varied between the 12 comparisons patients were asked to make. In each instance, patients were asked to

choose which hospital they would attend. Patients responses were analysed using choice-based conjoint analysis.

The authors found that, overall, surgery-specific information had the most impact on patient's choices, followed by the percentage of patients with a 'textbook' outcome and then a hospital's reputation. The atmosphere in the hospital and waiting time for surgery had the least impact. However, the impact of the surgery-specific information varied between surgical groups and had more impact in patients undergoing cholecystectomy than in those undergoing inguinal hernia repair and more impact in patients undergoing aorta reconstruction than in those undergoing oesophageal resection. The authors concluded that surgery-specific information was more important than general information when patients choose between hospitals. In terms of our theory under test, it suggests that patients find quality of care information more informative than more general information in supporting their decision about which hospital to choose. This is important because PROMs data are a much more specific indicator of the quality of a hospital for patients undergoing elective surgery than standardised mortality rates for cardiac surgery. This may reflect that surgery-specific information is a more accurate or representative indicator of the quality of care that patients will receive at a hospital for a particular operation than more general information.

Dixon *et al.*⁷⁹

This report detailed a large mixed-methods study that collected a range of data on patient and provider experiences and views of patient choice. Using surveys and interviews, the report examined patients' revealed preferences and their stated preferences, and also included a discrete choice experiment. Four local health economies were selected to provide a sampling frame and case studies, based on their potential for patient choice (high or low, as indicated by the number of providers within 60 minutes' travelling time) and the percentage of patients reporting having experienced choice (high or low) reported in a previous Department of Health-commissioned MORI patient choice survey, carried out in 2007. A survey was sent to a random sample of 5997 patients across the four case study areas, resulting in a response rate of 36%. The sample contained disproportionately more patients who attended independent sector treatment centres, which the authors compensated for by using weighting in their analysis of the questionnaire data.

The survey had two main components. The first asked patients a series of closed questions about their recent experience of patient choices, and the factors that informed those choices, to provide an assessment of patients' stated preferences. Patients were presented with a list of hospital attributes and asked to rate, on a 4-point Likert scale, how important they were in informing their choice. They were also asked to indicate which sources they had used to gain information about the performance of the hospital.

The second component consisted of a discrete choice experiment, where patients were presented with five elements of information about three fictional hospitals and asked to choose one hospital. The information, or attributes, consisted of:

- waiting time and journey time to the hospital, to represent information that could be obtained from the Choose and Book system
- data on performance ratings (number of cancelled operations, hospital infection rates, improvements in patients' health) and patients' views from surveys (friendliness of staff, communication with patients, cleanliness, facilities) to represent information that could be obtained from leaflets printed by PCTs and through the NHS Choices website
- previous experiences and the opinions of others, to capture 'softer' information about the hospital's reputation.

The level of each attribute for each hospital was varied in comparison with each other over a series of six questions, and patients were asked to choose a hospital in each instance. This enabled the authors to explore which patients were more likely to be willing to travel to a non-local hospital. Finally, patients were

also asked to provide information on their social and demographic characteristics and which hospital they attended, and, using these data, the authors were also able to examine patients' revealed preferences.

In addition, interviews were conducted with 19 patients, 49 senior members of staff from NHS and independent sector treatment centres and 25 GPs from across the four health economies.

The survey data enabled the authors to test the theory about patients being prepared to travel further for a higher-quality hospital if offered a choice. From their survey, the authors found that 49% of patients recalled being offered a choice. In response to a survey question, 73% of all patients were referred to their local hospital; of those who were offered a choice, 29% were referred to a non-local hospital, while 21% of those not offered a choice were referred to their local hospital. The authors also analysed the distances patients actually travelled to attend their appointments, and found that 39% of those not offered a choice and 53% of those offered a choice travelled beyond their nearest hospital. After weighting the data, the authors argued that, when offered a choice, between 5% and 14% more patients travelled beyond their local hospital. In terms of the theory under test, this suggests that most patients use their local hospital, but offering patients a choice slightly increases the chance that they will travel further than their local hospital for care. In their interviews, both GPs and providers perceived that patients were loyal to their local hospital because there was 'a strong feeling that people should support the local hospital' and because patients 'don't want to have to travel' (p. 62).⁷⁹

The authors examined which patients were more likely to choose a non-local hospital from the real choices they made. Analysing data from patients who indicated they had been offered a choice, using a binary logistic regression model with choice of a non-local hospital as the dependent variable, the authors found that respondents aged between 51 and 65 years were more likely than those aged 16–35 years to choose a non-local hospital; respondents who lived in villages or rural locations were more likely than those living in cities, large towns or suburbs to choose a non-local hospital; and those holding a degree were more likely than those with no formal qualifications to choose a non-local hospital. Of particular relevance to our theory under test, the authors found that those who had a bad or mixed previous experience with their local hospital were more likely to choose a non-local hospital than those who had a generally good previous experience.

The discrete choice experiment also provided further evidence of which patients were more likely to travel to a non-local hospital. The authors found that, across all choice scenarios, 25% of patients always chose the local hospital, irrespective of the hospital's characteristics, while 70% weighed up the different alternatives and sometimes chose a non-local hospital. The remaining 5% always opted for a non-local hospital across all choice scenarios. They found that patients without internet access, those with a low formal level of education, those who did not travel by car and those living in a big city were more likely to choose the local hospital. They found that those who would like more information to help choose their hospital, those with bad or mixed previous experience of their local hospital, those who had heard about the performance of hospitals in their area from the local media, those being referred for trauma or orthopaedics and those who had visited their GP six times or more in the last 12 months were more likely to choose a non-local hospital. In terms of our theories under test, both studies found that patients with a mixed or previous bad experience of their local hospital would be more likely to choose a non-local hospital, providing lateral support for the theory that if patients have a bad previous experience of a hospital they are more likely to 'exit' from a local hospital.

Findings from the survey indicated that, for patients who had been offered a choice, the five factors with the highest mean rating or importance (on a scale of 1–3) were hospital cleanliness, quality of care, standard of facilities, organisation of clinic and hospital reputation (*Figure 14*). Although patients rated cleanliness and quality of care as most important in their choice of hospital, in their interviews most GPs

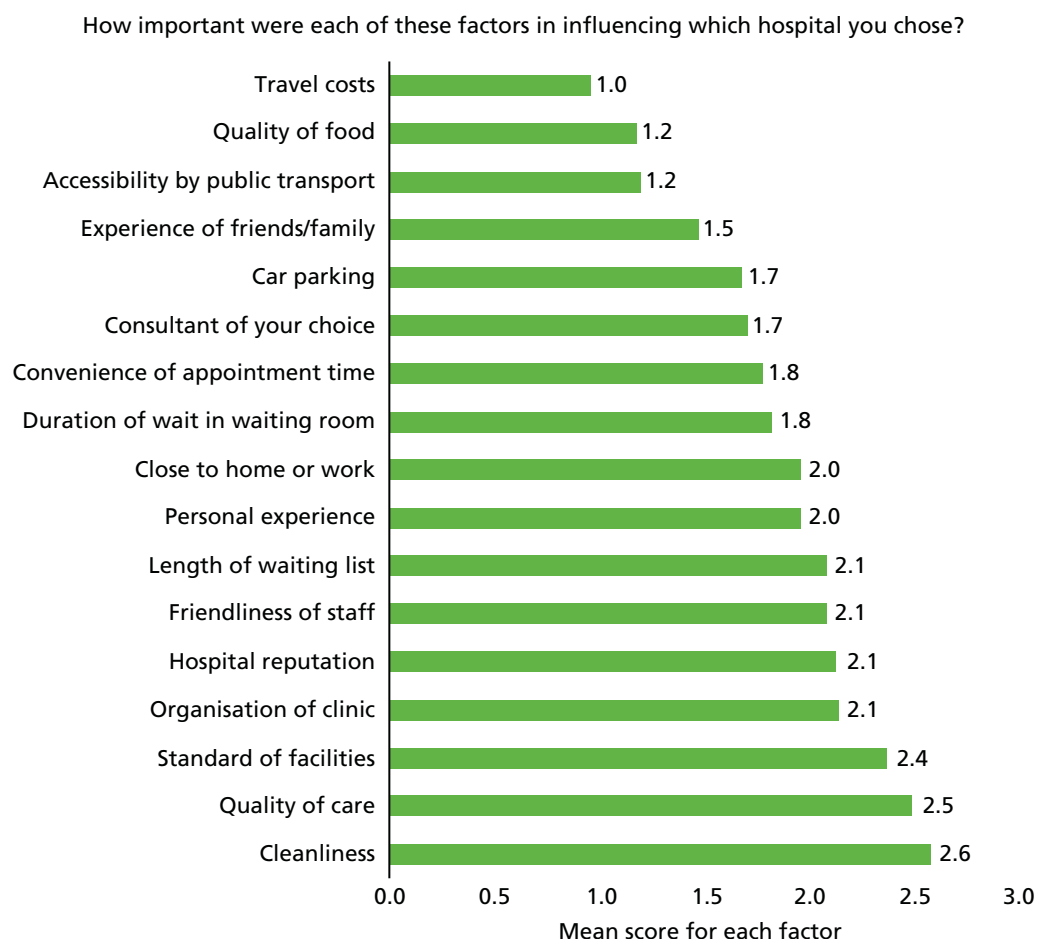


FIGURE 14 Factors that influenced patients' choice of hospital. Reproduced from Dixon *et al.*⁷⁹ with permission from The King's Fund. Note: in this figure, factors rated as 'essential' were scored as 3, 'very important' as 2, 'somewhat important' as 1 and 'not important' as 0. The scores shown are the mean score given by respondents for each factor. Scores for each factor have been rounded to one decimal place. The length of the bar represents the non-rounded score.

felt that patients did not judge the quality of a hospital based on clinical outcomes or performance ratings. As one GP indicated:

I don't think I've ever had a patient say 'I don't want to go there because I've looked at their outcome figure and I don't like the look of it'.

*GP, Area B. Reproduced from Dixon *et al.*⁷⁹ with permission from The King's Fund*

A hospital's reputation ranked fifth in patients' mean ratings of importance. In their interviews, GPs also perceived that this was an important factor for patients, and recalled instances in which patients had expressed a preference for or against a hospital based on its reputation. In their interviews, patients themselves also concurred that a hospital's reputation informed their choices. Their quotations suggest that patients form opinions about a hospital's reputation based on their own personal experiences, word of mouth and reports in the media:

One of the hospitals does have a bit of a reputation in the media and people do say 'Well, actually, I would rather go to [another hospital name] because of that'.

*GP, Area A. Reproduced from Dixon *et al.*⁷⁹ with permission from The King's Fund*

It's got a bit of a bad reputation lately. And I think things like that do put me off. [Is that from stories in the press or just generally what you hear locally?] Word of mouth, people I've known use the

hospital . . . And I think maybe things I've read in the paper, and, of course, you do read things like cleanliness and stuff like that, which, I think, straight away does put you off.

Female patient, aged 57, area B. Reproduced from Dixon et al.⁷⁹ with permission from The King's Fund

In terms of the theory under test, these findings suggest that patients rate 'cleanliness' and 'quality of care' as the most important factors that influence their choice of hospital. It is not clear from the report whether or how the term 'quality of care' was defined or how patients interpreted this term when completing the questionnaire. However, it appears that patients make judgements about these factors based on their perception of a hospital's reputation, which is informed by their experience, what they read in the press and word of mouth.

In the survey, patients were also asked to indicate whether or not they had received any advice about their choice of hospital. Of patients who had been offered a choice, 40% had received advice from their GP, 35% had consulted family and friends and 14% had received advice from a booking advisor. In their interviews, GPs reported that they tended to advise patients based on 'local knowledge' and 'feedback from previous referrals' (p. 87).⁷⁹ They also indicated that they tended to:

. . . stick to consultants we're happy with . . . to people that we know . . . [that] we've had a good positive experiences in the past.

Reproduced from Dixon et al.⁷⁹ with permission from The King's Fund

Patients were asked to indicate which sources of information they had used to choose their hospital. The most commonly cited source was their own experience (41%), followed by the GP (36%) and then friends and family members (18%). Six per cent of patients indicated that they had consulted a booklet or leaflet produced by the PCT, and 4% had consulted the NHS Choices website. Furthermore, patients were asked from which sources they had heard about the performance of the hospital. The most common sources cited by patients were personal experience (56%) and the experience of friends and family (52%), followed by the local media (28%), newspapers (22%) and the 'grapevine' or gossip (21%). In contrast to their role as a source of information to inform choice, GPs were much less commonly cited as a source of information on hospital performance (13%). Finally, 7% of patients had consulted official performance reports and 3% had consulted the internet. In their interviews, patients also reported that they 'don't look anything up on the internet' but preferred to 'base my opinions on my own experience, experience of close family' (p. 89).⁷⁹

This was a large and comprehensive study, which used a range of data to examine how patients make choices about hospitals. In terms of the theory under test, the findings suggest the majority of patients still go to their local hospital, but, if offered a choice, around 50% of those patients opt to travel further than their local hospital. Patients were more likely to choose a non-local hospital if they had had a bad previous experience with their local hospital, and indicated that 'cleanliness', the 'quality of care' and the reputation of the hospital were important considerations influencing their choice of hospital. However, patients formed opinions about these factors not through consulting hospital performance data but through their own experiences, the experiences of friends and family, word of mouth and, to a lesser extent, the media. GPs were a source of advice about the choice of hospital, but were much less cited as a source of information of hospital performance; GPs themselves tended to provide advice to patients based on their own relationships with and experience of a provider, rather than on performance information. Thus, the authors concluded that providers were motivated to 'maintain [their] reputation in order to ensure patients returned or, through word of mouth, spoke highly of their experience' and that 'reputation and loyalty, combined with patient's ability to choose not to return to a hospital, create pressure on providers to deliver a high quality service' (p. 159).⁷⁹ This suggests that providers are motivated to improve care in order to maintain their reputation and thus avoid patients going elsewhere. However, this study's findings suggest that performance data only contribute indirectly to influencing a provider's reputation, via what patients hear about the hospital in the newspapers or media. For the most part, a provider's reputation depends directly on the experiences of patients at the hospital, their families and the GPs who refer them.

Theory 6 summary

These studies suggest that fewer than half of patients recalled being offered a choice of hospital and, of those who did, a significant minority opted to go to a non-local hospital or to travel further than their local hospital. Patients were more likely to change hospitals if they had had a bad previous experience. They tend to base their decisions about choice of hospital on 'soft' information, such as their own experience, the opinions of friends and family and advice from their GP, rather than on publicly reported information on hospital quality, such as that found on the NHS Choices website. Similarly, GPs also provided advice to patients on hospital choice based on their personal experiences and knowledge of providers, rather than on formal quality information.

Chapter summary

In this chapter, we began with a basic, unrefined theory.

Theory 1

We found limited evidence^{173,174} of significant impact of PROMs feedback on patient care, but found evidence²⁷ that other forms of publicly reported performance data, in some circumstances, lead to improvements in patient care. However, these studies did not explain the circumstances in which, or process through which, the feedback of PROMs or performance data leads to improvements in patient care.

We then reviewed a large amount of evidence derived from studies utilising a wide range of different study designs to test and refine different theories about the mechanisms through which public reporting of hospital performance is thought to lead to improvements in patient care. These theories have proposed different ideas about how and why providers might take steps to improve the quality of patient care in response to the public reporting of performance. It was theorised that providers take steps to improve patient care because of:

- 'Intrinsic motivation': their professional ethos means that they are intrinsically motivated to maintain good patient care and will take steps to improve if feedback highlights a gap between their performance and expected standards of patient care.
- 'Market share': they feel threatened by the potential loss of market share that could occur if patients decided to choose alternative, higher-performing providers.
- 'Professional reputation': they wish to protect their professional or institutional reputation, which may have been damaged by being labelled as a poor performer in public.
- 'Competitive benchmarking': they are competitive and wish to be as good as or better than their peers.
- 'Collaborative benchmarking': they improve patient care through learning about and implementing the best practices of 'high-performing' organisations as a result of the sharing of information.

These theories are often presented as 'rival' explanations of how and why providers respond to performance data. However, our synthesis challenges this premise and suggests that, rather than being mutually exclusive, these different theories are interconnected. In this summary we bring the findings of our synthesis together to explain how they are interconnected, and offer a refined theory of how and why providers respond to the public reporting of performance data.

Theory 2

Evidence from a systematic review¹⁷⁶ and a study¹⁷⁷ examining private feedback suggest that it can motivate providers to respond through raising awareness of performance in relation to peers and through peer competition. Evidence from studies comparing private and public reporting^{35,39,118,179} suggests that while clinicians are intrinsically motivated to maintain good patient care, the public reporting of performance places additional pressure on them to take steps to improve patient care, particularly for poor performers. This pressure may derive from the threat of either loss of market share or damage to their reputation.

Theory 3

Evidence from studies of patients' revealed preferences indicates that although most patients do not change hospitals following public reporting of hospital quality,^{27,182,191} hospitals are more likely to experience a change in their market share if they have previously been a 'good' hospital and report cards provide new information that they are now a 'poor performer';¹⁸⁴ in other words, their reputation as a high performer is questioned. Here we see the first major revision to our initial theories. This suggests that the 'market share' and 'reputation' mechanisms are not separate, that damage to a providers' reputation can, in turn, lead to a reduction in their market share.

Theory 4

Evidence from surveys, interviews and ethnographic studies of providers' views and responses to the public reporting of hospital quality supports this idea and suggests that, although providers are more concerned about the consequences of public reporting on their reputation than on their market share,^{118,179,192,193} they also perceive that damage to their reputation can have a subsequent impact on whether or not patients choose to attend their hospital.^{18,194,195} One of the mechanisms through which providers perceive that report cards may damage their reputation is press or media coverage that either misrepresents the data or disproportionately focuses on reporting poor performance.^{18,195} They also perceive that such data do not provide an accurate indicator of hospital quality, and they are concerned that patients are not able to understand the data and may, therefore, make erroneous choices as a result.¹⁹⁵

Studies that examined newspaper reportage of hospital performance data indicated that newspaper headlines did not directly reflect the information contained in the articles, and that articles over-represented higher- and lower-performing hospitals while under-representing hospitals with average performance.¹⁹⁸ This suggests that newspaper reporting of hospital quality may have raised patients' awareness of hospital quality, but did not necessarily provide balanced information about it. Furthermore, evidence from patients' revealed preferences suggested that a hospital's market share was more influenced by 'tales of the unexpected' than by the public release of the report card itself or newspaper reports about it.¹⁸⁵ This suggests that patients' choice of hospital is much less influenced by report cards themselves and press reports about them than providers believe.

Theory 5

We also found evidence to support the theory that providers are motivated to take steps to improve hospital quality because they wish to be as good as or better than their peers.^{98,194,199} Furthermore, taking steps to improve the quality of care was perceived as enhancing a provider's reputation, which, in turn, meant that their peers were more likely to refer patients to that provider.

Theory 6

Finally, we tested this theory by reviewing studies that explored whether or not patients were offered a choice of hospital, whether or not they chose to attend a non-local hospital and the factors that influenced this. These studies, undertaken in England, indicated that < 50% of patients recalled being offered a choice of hospital⁷⁹ and, of these, 29% were referred to a non-local hospital and 53% travelled distances further than their local hospital.⁷⁹ Evidence from both patients' revealed preferences and a discrete choice experiment suggested that patients were more likely to choose a non-local hospital if they had had a bad previous experience with their local hospital, and indicated that 'cleanliness', the 'quality of care' (although it is unclear how this was defined) and the reputation of the hospital were important considerations influencing their choice of hospital.⁷⁹

However, several studies found that patients formed opinions about these factors not through consulting hospital performance data but through their own experiences, the experiences of friends and family, word of mouth and, to a lesser extent, the media.^{79,210,211} While GPs were a source of advice about the choice of hospital, they were much less cited as a source of information of hospital performance.⁷⁹ In line with this finding, GPs themselves tended to provide advice to patients based on their own relationships with and experience of a provider, rather than on the basis of performance information.⁷⁹ These studies lead us to

conclude that performance data do not influence patients choice of hospital or their opinion of its reputation; rather, it is their own experiences, those of friends and family and those of their GP that inform their choice.

Thus, our synthesis suggests a further major revision to our initial theories. Publicly reported information on hospital quality plays little role in patient's choice of hospital; patients do not use it to inform their choices and referrers do not use it to inform their choices or the advice they give to patients. The motivation to improve and provide good care is driven by a desire to provide patients with a good experience, so that they will tell their friends, family and GPs about their experiences, which in turn will maintain patients' loyalty to a hospital and enhance a providers' reputation and, therefore, patients will choose to go there and GPs will choose to refer them. Providers consider that publicly reported quality information can damage their reputation when it is misrepresented in the media or because patients may misunderstand it.

Conclusions

In conclusion, our synthesis started with the theory that patients will use publicly reported information on hospital quality to choose high-quality providers, which in turn will lead to providers to fear a loss to their market share and they will respond by take steps to improve patient care. An alternative theory was that providers take steps to improve patient care because they fear the public report of hospital quality will damage their reputation. A further theory was that providers will take steps to improve the quality of patient care in order to be as good as or better than their peers. Our synthesis suggested a number of revisions to these theories. First, it suggests that patients do not use publicly reported quality information to inform their choice of hospital, but instead rely on their personal experience, the opinions of friends and family and advice from their GP to make the choice. Second, it suggests that these theories are not mutually exclusive but are interconnected; providers seek to improve the quality of care in order to maintain their reputation and maintain their market share. Providers take steps to improve the quality of patient care in order to provide patients with a good experience of care and maintain their loyalty, so that the patient will tell their family and friends, who will then believe that the hospital has a good reputation and will, therefore, be more likely to also choose to go to the hospital. Providers also take steps to be as good as or better than their peers, which in turn enhances their reputation and means that peers are also more likely to refer patients to them. Publicly reported quality information itself plays little or no role in these processes; it is only perceived as a threat if it is misrepresented or misreported in the media, when it can, in the eyes of providers, damage their reputation.

Chapter 5 Feedback of aggregate patient-reported outcome measures and performance data: reviewing contexts

The previous chapter considered the mechanisms through which the feedback and public reporting of performance data might work. In this chapter, we review some of the circumstances or contexts that shape which of these mechanisms are triggered and thus how the feedback and public reporting of PROMs and other performance data may (or may not) improve patient care. It is important to note here that we are not simply analysing single contextual constraints. Programmes never operate in isolation, and the feedback of performance data has been inserted into complex health systems in which a range of concomitant innovations, policy initiatives and management directives also operate, which may sharpen or blunt the intended impact of PROMs and performance data feedback. *Figure 15* illustrates how an intended outcome of a programme will often distort in further contexts subject to further policy measures. Therefore, it is more appropriate to consider these as contextual configurations.

In this chapter, we explore how contextual configurations may trigger some of the intended and unintended consequences of feedback and public reporting. To recap, the intended consequence of public reporting is that clinicians take steps to improve the quality of patient care. The unintended consequences of feedback and public reporting of performance are that clinicians may:

- dismiss or ignore the data
- engage in 'effort substitution or tunnel vision' (i.e. focusing on the areas of care measured by the performance data to the detriment of other important, but unmeasured, areas of care)
- engage in 'gaming' the data (e.g. the manipulation of data to give the impression of change without any real change in the underlying performance).

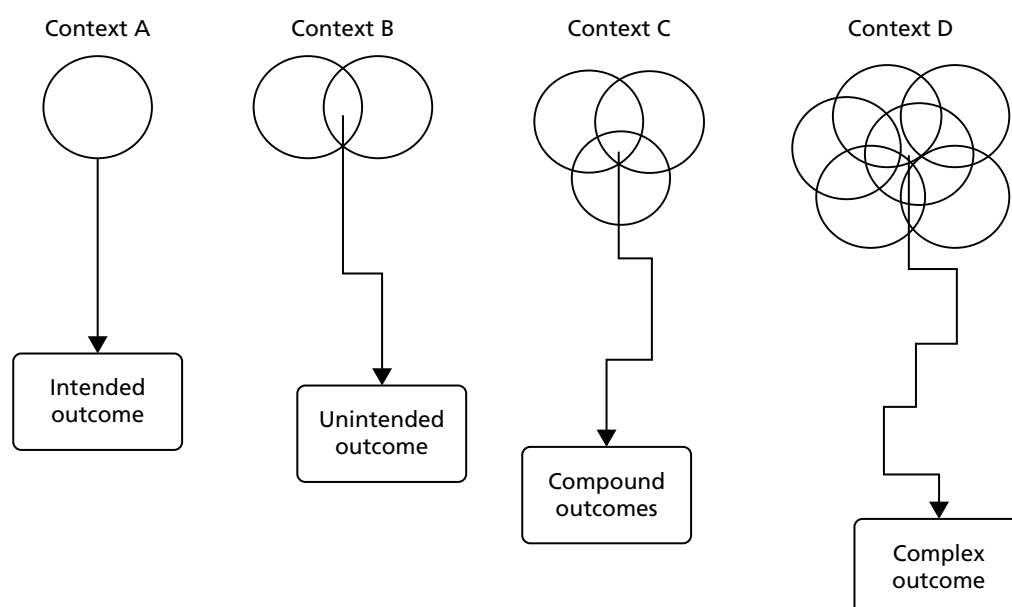


FIGURE 15 Contextual complexity and outcomes.

The contextual configurations that may influence how providers respond to the feedback and public report of performance data include:

- whether any rewards or sanctions were attached to performance
- the perceived credibility and validity of performance data
- the 'action-ability' of performance data.

Theory 7: financial incentives and sanctions influence providers' responses to the public reporting of performance data

Both benchmarking and public disclosure theories hypothesise that the power relationships and relative status of the organisations producing and acting on the performance data can influence the response of organisations whose performance is being assessed. This power may be exerted in a number of ways, for example through increased scrutiny, varying the relative freedom offered to organisations, and financial incentives and sanctions. A number of systematic reviews have highlighted the variable impact of financial incentives on professional behaviour.^{213,214} Pawson's programme theory of public disclosure programmes also recognises that publicising performance data rarely works in isolation from other sanctions. His realist synthesis²¹ of public disclosure interventions across a range of different contexts found that public disclosure is more likely to achieve its intended outcomes when it is targeted at aspirational elites and can be dovetailed with existing market sanctions.²¹ It is evident that previous and existing public reporting programmes are often accompanied by a range of incentives and sanctions. For example, in England, under the 'hospital star ratings' system, trusts achieving a three-star rating were granted 'earned autonomy' in the form of less frequent monitoring and inspections from the CHI, retention of profits from the sale of hospital land to reinvest in services and the right to become a foundation trust. Their ratings also determined the level of discretion chief executives had to make use of the 'NHS Performance Fund' to incentivise QI at a local level. Trusts with a zero-star rating were required to produce a 'Performance Action Plan' indicating the steps taken to improve care, which had to be agreed with the Modernisation Agency and the trust's Department of Health regional office.

Currently, PROMs data may form part of the indicators used to reward providers as part of the CQUIN system. From 2014–15, PROMs have been included in the BPT for hip and knee replacement. Providers would qualify for the BPT if they met the following criteria:

- do not have an average health gain significantly below the national average (99.8% significance), and adhere to the following data submission standards:
 - have a minimum PROMs pre-operative participation rate of 50%
 - have a minimum NJR compliance rate of 75%
 - have a NJR unknown consent rate of < 25%.

For 2015–16, NHS England proposed that the threshold for the NJR compliance rate be increased to 85%. Between 2004 and 2013, the QOF rewarded GPs for using standardised depression measures for screening and then reassessing patients suspected of having depression.

In this section we test the theory that attaching financial incentives to the public reporting of quality may accelerative and amplify the impact of public reporting and feedback of performance data on improvements to patient care. We also consider the theory that they may have a detrimental impact on aspects of care that are not incentivised or lead to the gaming or manipulation of data. We begin by considering a collection of quantitative studies that have compared the impact of public reporting alone with the impact of public reporting and financial incentives on the quality of patient care.

Theory 7a: financial incentives accelerate and amplify the impact of public reporting and feedback of performance data on improvements to patient care

Lindenauer et al.²¹⁵

This study makes use of a 'natural experiment', in which changes in hospital performance measured by quality indicators for hospitals that voluntarily participated in public reporting of performance scheme are compared with the performance of a subset of these hospitals who also voluntarily participated in a pay for performance scheme in the USA. The public report scheme, the Hospital Quality Alliance (HQA), was initiated in 2002, with all acute hospitals in the USA invited to participate and incentivised to do so by linking participation with the annual Medicare payment update; 98% of hospitals participated in the scheme. They were expected to collect and report on a minimum of 10 quality measures across three conditions: heart failure, myocardial infarction and pneumonia.

The pay-for-performance scheme in the Centers for Medicare and Medicaid Services (CMS)-Premier Hospital Quality Incentive Demonstration (HQID) was initiated in 2003. A total of 421 hospitals who subscribed to a quality benchmarking database were also invited to participate; 266 agreed but 11 later withdrew, leaving 255 hospitals. As part of this programme, hospitals were expected to collect and publicly report on 33 quality measures for five clinical conditions (heart failure, myocardial infarction, pneumonia, coronary bypass grafting, and hip and knee replacement), which also included the 10 indicators as part of the HQA programme. In addition, for each clinical condition, hospitals in the top decile on a composite measure of quality for a given year received a 2% bonus payment, while hospitals in the second decile received a 1% bonus payment. Hospitals that, at the end of the third year of the programme, had failed to exceed the baseline performance of hospitals in the lowest two deciles incurred financial penalties of between 1% and 2%.

The authors included hospitals if they submitted data on a minimum of 30 cases for a single condition annually as part of the HQA programme. They matched the 255 hospitals that participated in the HQID programme with at least one hospital that participated in the HQA programme alone on the basis of number of beds, teaching status, region, location and ownership status. They matched 199 HQID hospitals each with two HQA hospitals and eight with only one; thus, a total of 207 HQID and 406 HQA only hospitals were included. They compared the change in adherence to performance on 10 indicators shared by both programmes over eight quarters for each hospital between 2003 and 2005, and also calculated change in adherence to two compound measures for each of the three conditions (myocardial infarction, heart failure and pneumonia). They also recalculated the differences adjusting for confounding variables using a linear regression model.

The authors found that pay-for-performance hospitals showed statistically significantly greater improvement for 7 out of the 10 individual measures and in all of the composite measure scores. When differences in baseline performance and other confounding variables were taken in to account, the incremental effect of financial incentives decreased from 4.3% to 2.6% for the composite measure for myocardial infarction, from 5.2% to 4.1% for heart failure and from 4.1% to 3.1% for pneumonia; all differences remained statistically significant. The authors concluded that this suggests that the financial incentives have a modest effect on 'catalysing quality improvement efforts among hospitals already engaged in public reporting'. However, some caution must be exercised in interpreting these results. Those who participated in the HQID programme were likely to be enthusiasts, and it is likely that they were the better performers at baseline and, thus, more likely to continue to improve their performance. It is unlikely that matching or statistical control of confounders accounted for all the possible differences between the HQA and HQID hospitals. Furthermore, the study assessed the HQID hospitals performance on only 10 indicators but they publicly reported on 33. However, in terms of the theories under test, this study provides some evidence that financial incentives and sanctions can accelerate or amplify provider responses to public reporting programmes.

Friedberg et al.²¹⁶

This study examined doctors' groups' responses to and use of the publicly reported patient experience data, and compared the characteristics of groups that had different levels of 'engagement' or use of the data.

The data were collected by the Massachusetts Health Quality Partners (MHQP) collaborative, which had been publicly reporting patient experience data since 2006. A total of 117 doctor groups were invited to participate in a 30-minute semistructured interview and 72 (62%) of group leaders responded. The interviews explored group leaders' use of patient experience reports, what sort of improvement activities had been initiated as a result and data on the characteristics of the group (e.g. group size, organisational model, employment of doctors and exposure to financial incentives).

The initial step of their analysis identified three different levels of doctor group engagement with patient experience data: level 1 groups did not recall receiving patient experience surveys and did not use them, apart from distributing the reports to their staff (17% of their sample); level 2 groups took one or more actions to improve quality, but these were largely directed at doctors or sites that were low performers (22%); and level 3 groups reported one or more group-wide initiatives to improve patient experience, which included most or all staff or sites in the group. They found that level 3 groups were statistically significantly more likely to be integrated medical groups, to employ their own doctors, to be network affiliated and to be exposed to financial incentives based on measures of clinical quality. The authors concluded that their findings indicated that the improvement strategies require a 'managerial infrastructure capable of starting and directing improvement activities' improved by 'payment incentives based on patient experience'. Some caution must be exercised in interpreting these results, as the authors' findings were based on self-report; respondents may have over-reported their QI activities; and there may have been other, unexplored factors that may have explained differences in the level of engagement in doctors' groups. Nonetheless, in terms of our theory under test, the results provide some evidence to support the idea that financial incentives can increase the likelihood that providers will initiate QI activities in response to the public reporting of performance.

Alexander *et al.*²¹⁷

This study examined the impact of public reporting and financial incentives on the extent to which small and medium-sized doctor practices in the USA engage in 'care management practices' (CMPs). The authors define CMPs as 'organized processes implemented by doctor groups to systematically improve the quality of care for patients'.²¹⁷ These include the use of patient registers, electronic medical records, doctor performance feedback and provider education. As such, the study does not examine the impact of public reporting and incentives on the quality of patient care per se, but on processes that are thought to lead to improvements in the quality of patient care.

Their analysis is based on survey data collected as part of the national study of small and medium-sized physician practices; these are practices with < 20 practising doctors. This survey focused on the 14 'communities who were in receipt of support from the Aligning Forces for Quality programme'. Practices that were part of the Aligning Forces for Quality programme were provided with both grants and support from people with expertise in QI to support them to both measure and publicly report on the quality of care, and to take steps to improve care and involve patients in this process. The questionnaire was sent to a stratified random sample of the Aligning Forces for Quality practices ($n = 1793$), of whom 67% responded ($n = 1201$). The authors of this paper focused on a subsample of 643 practices who were engaged in either private or public reporting of quality. They used the Physician Organisation of Care Management Index (PCOMI) as an indicator of the level of CMP use. The PCOMI is a summary measure (ranging from 0 to 24) of whether the practice uses reminders for preventative care, uses doctor feedback, has a disease register, has clinical practice guidelines and employs non-doctor staff educators in the care of people with four chronic conditions (heart failure, depression, asthma and diabetes). To explain variation in the PCOMI, the authors constructed binary indicators of (1) whether the quality performance of the practice was publicly reported; (2) whether the practice received a financial reward on the basis of its quality performance during the past year; and (3) whether the practice was aware of quality reports.

The authors found that, controlling for patient and practice characteristics, practices that received financial rewards engaged in a statistically significantly higher number of CMPs, with a PCOMI score 7 points higher than those practices that were not in receipt of financial rewards. They also found that practices that

discussed quality reports at their physician meetings also engaged in a statistically significantly higher number of CMP practices, with a PCOMI score 17 points higher than those of practices who did not discuss quality reports. Practices that were subject to public reporting did engage in a higher number of CMP processes than practices that were not subject to public reporting, with PCOMI scores 8 points higher than those of practices not subject to public reporting; however, this difference was not statistically significant. Finally, they found that practices who were subject to both public reporting and financial rewards engaged in a higher number of CMPs, with PCOMI scores 10 points higher than those of practices subject to only public reporting or financial rewards. The authors argued that their findings demonstrate there is a 'significant joint effect of having both PR [public reporting] and financial incentives above and beyond just having one of them'.²¹⁷ In their discussion, they acknowledged that their findings may indicate that doctors who participate in the public reporting of quality require financial rewards in order to invest the time required to change clinical care, or that those practices which have CMPs already in place are more likely to produce quality outcomes worthy of financial reward. Their study did not allow for an exploration of doctors' motives.

This was a cross-sectional study and, therefore, it is not possible to infer a causal relationship between financial incentives, public report and CMPs. We can also question the validity of the PCOMI as a measure of provider engagement in CMP practices. However, in terms of the theory under test, this study provides a further layer of evidence supporting the idea that financial incentives may amplify the impact of public reporting on the quality of patient care. The study also suggests that whether or not practices were aware of and discussed the public reports of quality was more important than whether or not they were subjected to public reporting.

Doran *et al.*²¹⁸

This paper reported on a longitudinal analysis to compare achievement rates for 23 activities included in the English primary care QOF incentive scheme and 19 activities not included. The achievement of the QOF indicators was also publicly reported through the HSCIC website and on the NHS Choices website. As such, the study allows for an exploration of the theory that financial incentive schemes can lead to the neglect of activities that are not included in the scheme, the so-called 'tunnel vision' hypothesis. The authors' analyses drew on data from the General Practice Research Database, which contains patient data on morbidity, treatment, prescribing and referral from 500 general practices and covers 7% of the UK population. A sample of 148 practices that provided data continuously throughout the study period (2000–7) was selected to reflect practices with a range of list sizes. A random sample of 4500 patients was drawn from each practice.

The 42 indicators were selected from a pool of 428 indicators identified by the research as being already established indicators or those based on clinical consensus, as expressed in national guidelines. They excluded indicators that may have been affected by significant changes in the underlying evidence base or were dropped from the QOF scheme in an attempt to rule out some of the possible effects of other changes in the policy environment on achievement of the indicators. They classified the indicators into two subtypes: those relating to measurement (e.g. blood pressure measurement) and those relating to prescribing, as previous research suggested that response to these types of indicators may be different. Thus, the indicators were classified into four groups by type and whether or not they were incentivised.

The difference between the expected achievement rate and the actual achievement rate for the indicators was analysed using multivariate regression models over four different time periods: before the introduction of the QOF scheme (2000–3); during preparation for the QOF – when practices knew the scheme would be introduced but did not yet know details of the indicators (2003–4); immediately after the introduction of the QOF (2004–5); and longer term (2005–7). The authors examined the impact of the incentive scheme on the four indicator groups separately and then compared incentivised and non-incentivised indicators for each of the two types (measurement and prescribing).

The authors found that, prior to the introduction of the QOF (2000–3), achievement increased significantly for 32 out of the 42 indicators, decreased significantly for two and did not change for eight. The authors argue that these findings suggest that quality initiatives introduced over this period such as clinical audit, and the introduction of statutory bodies focused on QI (such as NICE), had an impact of improving the quality of care in general practice. The authors found that achievement rates improved at the fastest rate *prior* to the introduction of the QOF (between 2000 and 2003) for those indicators that were subsequently incentivised under the QOF system. Thus, the QOF indicators focused on areas of care that previously had already shown the greatest improvements, suggesting that they reflected areas of care in which practices were already performing well and/or perceived to be important.

In the first year following the introduction of the QOF, achievement rates for incentivised indicators increased substantially for all measurement indicators, with increases above the predicted rates in 2004–5 of up to 38%. The prescribing indicators had a higher baseline rate than the measurement indicators, increased at a slower rate during the pre-intervention period and, although they also saw significant increases above the predicted rates in the first year, these increases were smaller than those found for the measurement indicators (1.2–8.3%). However, collectively, the incentivised indicators ‘reached a plateau in the second and third years of the scheme’ where ‘only 14 of the 23 incentivised indicators had achievement rates significantly higher than rates projected from pre-intervention trends after three years’.

For the non-incentivised indicators, achievements rates immediately following the QOF improved in line with achievement rates projected from pre-intervention rates. However, in the second and third years, the rate of QI slowed down relative to expected achievement rates. By 2006–7 the authors found that quality was significantly worse than expected from pre-intervention trends, especially for measurement activities. Improvement rates were also significantly lower, relative to projected rates, than achievement rates for the incentivised indicators.

This study did have a number of limitations, acknowledged by the authors, that may have affected its conclusions. Achievement rates may have been affected by changes in the consistency and accuracy of recording indicators over time, especially for incentivised indicators. They may also have been influenced by changes to the case mix of patients subject to the indicators, especially as the incentivised indicators also encouraged increased case finding (e.g. for depression). The practices selected may not have been representative of the population, although the trends found in this study replicated those seen nationally. The authors focused on a limited selection of indicators, which may limit the generalisability of their findings, but at the same time increased the likelihood of attributing differences in achievement rates to the incentivised/non-incentivised status of the indicators, rather than due to other contextual changes. As such, we can be reasonably convinced that this study provides a useful test of the theory that financial incentives may lead practitioners to focus on incentivised aspects of care, to the detriment of non-incentivised aspects.

This study found that although financial incentives increased the quality of care in the short term, in the longer term their impact petered out. The authors suggest three possible explanations for their findings: (1) that the improvements seen in the first year were due to better recording procedures; (2) that practices reached a ‘ceiling’ limit of quality in the third year, with little opportunity for further improvement; or (3) that practices took their foot off the accelerator for these incentivised activities because they had already reached the threshold at which they would receive the maximum amount of remuneration, and further effort would not result in increased income. Of particular interest to our synthesis is the finding that there was no detrimental impact on aspects of care that were not incentivised in the short term, but in the longer term the scheme had ‘some detrimental effects’ on certain areas of care, particularly measurement activities. In terms of the theory under test, this provides some support for the idea that financial incentives can lead to providers focusing on aspects of care that are incentivised at the expense of other areas of care.

Theory 7a summary

The studies reviewed in this section provide some evidence that financial incentives, together with public reporting, have a greater impact on improvements to the quality of patient care than either initiative alone. However, they also suggest that this impact may occur only in the short term, and that in the long term the impact of financial incentives may reduce, especially if providers reach the threshold at which they would receive the maximum amount of remuneration. Furthermore, they also provide some evidence to suggest that financial incentives may also lead to a 'tunnel vision' effect, whereby providers focus on incentivised aspects of care at the expense of non-incentivised aspects. The cross-sectional nature of most of the studies reviewed means that inferences of a causal link between public reporting combined with financial incentives and improvements in patient care need to be treated with caution. Furthermore, the studies reviewed above do not provide any insights into how providers themselves have responded to public reporting when financial incentives were attached. We now consider a series of qualitative studies that have examined how providers have responded to either public or private reporting of quality both when financial incentives are attached and when they are not.

Theory 7b: providers do not make improvements to patient care when no financial incentives are attached to performance feedback

We start by reviewing studies examining responses to private and public reporting of quality when no financial incentives are attached to performance. Here, we test the theory that providers do not make improvements to patient care when no financial incentives are attached to performance feedback.

Wilkinson *et al.*²¹⁹

Wilkinson *et al.*²¹⁹ explored the views and responses of 52 staff, in 15 general practices, to cardiovascular and stroke performance indicators that were developed by the authors. The indicators largely focused on process and were similar to those later included in the National Service Frameworks for coronary heart disease and later the QOF. However, this study was undertaken before those initiatives were implemented. For half of the practices, the academic team collected the data themselves, and for all practices they fed back data to the practice in a 1-hour presentation; as such, the indicators were fed back *privately*, rather than publicly. During this presentation, the academic team explained how the indicators were developed, how the indicators for that practice compared with those of other practices in the study and the potential clinical benefits if full uptake of the indicators were obtained by the practice. The practice was encouraged to develop an action plan to address changes that they practice felt were necessary. Two months after the presentation, the authors interviewed a range of staff at each practice, including the GPs who led audit activities within the practice ($n = 15$), other GPs ($n = 14$), practice nurses ($n = 12$) and practice managers ($n = 11$).

The authors found that almost all of the GPs and nurses, and half of the practice managers, questioned the validity of the data used to generate the indicators because of gaps in the data, computer-related difficulties and confusion in applying Read codes. The most common response to the feedback of the indicators was to improve the number, uniformity and accuracy of recording data, with almost half of the practices attempting to do this. The most common reason cited was to 'demonstrate to other practices within the primary care group that their own practice was providing good care'²¹⁹ and, less commonly, to prompt GPs to improve patient care. Three of the 15 practices initiated an audit to validate the data used to produce the indicators. The authors note that 'all the professionals found the comparative nature of the results useful in interpreting their practice's performance'.²¹⁹ In their interviews, respondents mentioned that the comparative indicators had highlighted a gaps in their own performance compared with their own perceptions, and differences between their own performance and that of their peers:

[O]ne imagines that one is doing a fantastic job, then when you actually see it in writing you think oh that's not quite as good as you think. I am sure that this sort of presentation really winds you up to do better.

GP

It is helpful to be able to compare to local means and see whether you are doing a bit better or worse, and that perhaps is one of the strongest ways of getting GPs to alter things . . . they like to be seen to doing things a bit better than their colleagues.

GP. Reproduced from Quality in Health Care, Wilkinson EK, McColl A, Exworthy M, Roderick P, Smith H, Moore M, Gabbay J, vol. 9, pp. 166–74, © 2000, with permission from BMJ Publishing Group Ltd²¹⁹

Out of the 15 practices, 11 developed action plans for change, but these were largely in the form of ‘informal verbal agreements devised by one or two enthusiasts who were usually doctors’ and most focused on single changes. Communication of the action plans within practices was ‘ad hoc and informal’.²¹⁹ The authors identified that change was constrained by a lack of time and resources (both financial and human) to act on the information. Change was supported if the indicator represented a personal interest of one of the practice members or there was someone in the practice who had been allocated responsibility for that clinical area. Change was also supported when the indicators ‘accorded with other ongoing local and national initiatives’ which serve to increase ‘the status or relevance of the indicator results’.²¹⁹

In terms of the theories under test, this study indicates that, in line with feedback and benchmarking theories, the private feedback of their performance to GP practices highlighted gaps between their own performance and the indicators, and highlighted differences between the performance of their practice and that of other practices. This prompted the practices to reflect on their own practice, both as individuals and as an organisation. However, it also suggests that GPs had little trust in the validity of the indicators and sought to improve the accuracy of their data, or, less commonly, to investigate the reasons behind the indicator findings via audit. We could hypothesise that these activities were important to (a) improve the trust the practices had in their data and (b) provide a further basis on which to initiate any changes. However, the authors found that, beyond these activities, practices made few formal, co-ordinated attempts to change practice because they did not have the time, resources or interest in doing so. The authors argued that the ‘absence of specific incentives to change, either positive or punitive, meant that responses were purely voluntary’.²¹⁹ In terms of the theories under test, this suggests that the private feedback of data that are not trusted by their recipients, and without incentives attached to them, may prompt individuals or organisations to reflect on their practice and to improve the accuracy of the data collected, but it does not lead to any longer-term improvements to patient care.

Mannion and Goddard^{220,221}

These papers report on provider and health board responses to the Clinical Resource and Audit Group (CRAG) indicators in Scotland. The CRAG indicators were compiled and disseminated by the then Scottish Executive. They consist of seven reports including 38 clinical indicators covering a range of specialties, and individual trusts and health boards are named in the reports. They were not part of a formal framework of performance management, and the Scottish Executive indicated that they should not be used to make definitive judgements about the quality of services.

The authors conducted case studies of eight Scottish NHS trusts that varied by size, geographical area and performance on the indicators. They focused on the impact of two specific indicators relating to 5-year survival of women with breast cancer and 30-day survival after admission for stroke. All eight trusts were ‘average’ for the breast cancer indicators; six of the eight were average for the stroke indicator and two were worse than average. In each trust, they interviewed the chief executive, medical directors, consultants with responsibility for stroke services, consultants with a responsibility for breast cancer services, nurse managers and junior doctors ($n = 48$). The authors also interviewed the director of public health (or deputy) at the local health board for each trust, and key staff from the Information and Statistics Division of the Scottish Executive and the CRAG secretariat, to explore the intended purpose of the indicators.

With the trusts, the authors found that the CRAG indicators were ‘rarely cited by staff as the primary driver of QI or sharing best practice between organisations’.²²⁰ The indicators were not integrated into formal

clinical governance systems, but were mainly used by trusts to argue for increased resources for services. For example, for the two trusts that were 'worse' than average for the stroke indicators, the CRAG data were used to argue (successfully in one case) for a new stroke unit in order to improve care. These trusts also conducted further audits to check the quality of the data, and one introduced patient protocols. Two of the 'average' trusts also used the CRAG data to argue for a new stroke unit, while, in another, the health board declined a request for additional funding because the CRAG data were satisfactory. For three other 'average' trusts, the CRAG data had no discernible impact on stroke services. In response to the breast cancer indicators, in three trusts the data were used alongside other information to inform or argue for the set-up of new services.

Similarly, they found that CRAG indicators were rarely used by health boards to make definitive judgements about the quality of care in trusts. If they were used at all, they were used to highlight potential problems requiring further scrutiny, especially if the trust was identified as a significant outlier. The health board would then meet senior staff within the trust to 'express concern and explore the problem in further detail'.²²¹

The authors identified some of the reasons for the low impact of the indicators. Both health boards and trust staff, especially consultants, questioned the quality of the data (e.g. inconsistent coding and quality of case-mix adjustment) and did not perceive them as credible. The data were also not perceived to be timely, due to the time lapse (at least 1 year) between their collection and publication. The data also appeared to be poorly disseminated to frontline trust staff; consultants and chief executives were aware of the data, but nurse managers and junior doctors were not. Similarly, within health boards, data were discussed at board level and by senior management, but were disseminated downwards only if 'the Health Board or a specific speciality was identified as being a significant national outlier on the indicators'.²²¹ The indicators were not part of a formal system of performance assessment and, as such, there were 'weak formal incentives for staff to perform satisfactorily on the indicators'.²²¹ Furthermore, health boards 'did not hold staff accountable for their performance'²²² and they were not used as a basis for changing contracts to a different trust. However, some trust staff did acknowledge that the indicators could sometimes enhance their 'professional status and reputation'.²²¹

These findings suggest that trust and health board staff did not perceive the CRAG indicators to be credible, and that the indicators were poorly disseminated within these organisations. Both good and poor performance on the indicators was used to justify requests for additional resources but performance on the indicators was rarely used as a basis to initiate improvements patient care. In terms of the theories under test, the CRAG indicators were not linked to any rewards and sanctions, and in their discussion the authors observed that 'Many Trust and Health Board staff identified this as the key reason why the indicators effected little change in provider organizations'.²²¹ However, they also noted that 'the introduction of explicit incentives may lead to reduced performance if this crowds out intrinsic professional motivation'.²²¹

Theory 7b summary

These studies suggest that when no financial incentives are attached to the public or private reporting of performance, and when stakeholders do not perceive data to be credible, they are rarely used as the basis to initiate improvements in patient care. The studies also suggest that, under these conditions, providers' first response to indicators is to verify the data on which the indicators are based, to improve the quality of the data or carry out audits. These activities may serve to (a) improve trust in the data and (b) provide a further basis on which to initiate any changes. However, these studies also suggest that, without the co-ordination, resources or incentives to do so, these preliminary investigations may not lead to longer-lasting change, and quality reports are largely ignored.

Theory 7c: financial incentives attached to the feedback of performance data can lead to 'tunnel vision'

The studies reviewed in theory 7a suggest that financial incentives can lead to providers focusing on aspects of care that are incentivised at the expense of other areas of care, or 'tunnel vision'. We test this

theory by considering a series of studies that have examined how providers respond to performance data when a specific set of incentives and sanctions are attached to performance.

Mannion *et al.*¹⁸

This study examined the impact of the NHS hospital 'star ratings' on acute hospital trusts in England. Recall from *Chapter 3* that the star ratings were a single summary score of hospital performance based on their achievement according to a range of indicators that were made publicly available. This system was also accompanied by a combination of financial rewards and sanctions. Trusts achieving a three-star rating were granted 'earned autonomy' in the form of less frequent monitoring and inspections for the CHI, retention of profits from the sale of hospital land to reinvest in services and the right to become a foundation trust. Their ratings also determined the level of discretion chief executives had to make use of the 'NHS Performance Fund' to incentivise QI at a local level. Trusts with a zero-star rating were required to produce a 'Performance Action Plan' indicating the steps taken to improve care, which had to be agreed with the Modernisation Agency and the trust's Department of Health regional office. The authors used a multiple case study design with purposeful sampling of high-performing ($n = 2$) and low-performing ($n = 4$) trusts based on 2000–1 performance data. They undertook documentary analysis (CHI reports and internal governance reports) and semistructured interviews with between 8 and 12 key managers and senior clinicians in each site. As such, their interview findings reflect the views of senior, rather than frontline, staff.

Participants expressed a general view that star ratings did not adequately reflect hospital performance, in terms of either their coverage or their sensitivity to local factors that were perceived to be beyond their control. Low-performing trusts especially felt that areas of excellent practice were not taken into account in the indicators. The study also reported that the star ratings had served to align internal performance management activities with national targets and direct resources on those aspects of performance seen as important by government. Some staff in low-performing trusts reported that star ratings were useful in illuminating dysfunctional senior management that previously remained unchallenged.

However, this study also found a number of unintended consequences of the star ratings. Some trusts reported manipulating and misrepresenting the data (e.g. not accurately reporting the number of 12-hour trolley waits) and gaming the data (e.g. cancelling operations the night before rather than on the day) to improve their ratings. These findings align with the hypotheses from benchmarking and public disclosure theories, that the lower the acceptance of the data, the more likely organisations are to engage in efforts to improve the presentation or appearance of the indicator. Another unintended consequence reported in this study was a perception that public disclosure had led to 'tunnel vision', with trusts focusing on the issues measured to the exclusion of unmeasured but important areas. One example reported in the paper was that the waiting time target of 13 weeks in children's services had 'forced the trust to concentrate on children referred to it by doctors, rather than professionals, even though the clinical needs of the patients may be very similar'.¹⁸

In their discussion, the authors concluded that hospitals use public reports as 'a lever to influence staff behaviour' and noted that the 'unintended and negative consequences of the star rating system came across loud and clear'.¹⁸ The authors contrasted the high profile of the English star ratings with the relatively low profile of the Scottish CRAG indicators. They hypothesised that one reason for this was the effectiveness of the dissemination strategy and the simplicity of the rating system, making it easier for both professionals and the public to understand. However, a consequence of this simplicity was that few participants in this study felt that the star ratings adequately reflected the quality of the hospital and dismissed the ratings as invalid. We note here that a further key difference was that star ratings had a system of rewards and sanctions attached to them, whereas the CRAG indicators did not. In terms of the theory under test, this study suggests that when rewards and sanctions are attached to public reporting, but providers do not accept the validity of indicators, their efforts may focus on demonstrating the appearance of a high performance, resulting in unintended consequences such as effort substitution and gaming.

Dowrick *et al.*⁸¹

This study aimed to examine both GPs' and patients' views about the introduction of the routine collection of standardised measures of depression severity, which was incentivised in the QOF. Under this framework, GPs received QOF points for administering a standardised measure of depression for patients coded as being newly diagnosed with depression, and also received points for reassessing patients using a standardised measure of depression 2 weeks later. Furthermore, GP QOF scores were made available to the public via the HSCIC website from 2005 onwards. The authors interviewed 34 GPs and 24 patients from 38 practices in three locations in England. Here, we focus specifically on GPs' views of the validity of the standardised depression measures and how the incentives built into the QOF influenced GPs' use of the measures in practice.

The authors found that GPs questioned the validity of standardised measures of depression, both as clinical tools to aid patient management and as aggregate measures of the prevalence of depression. For example, one GP indicated, 'I don't have sufficient confidence that it's an objective enough tool, really, to measure trends'.⁸¹ They also expressed scepticism about the perceived motivation behind this QOF indicator, suggesting that its introduction was based not on an extensive consideration of the evidence that the indicator would improve care but on the hunches of a few academics and policy-makers to make extra work for GPs. As one GP remarked, 'I have a horrible feeling that a few academics got together and said this is a good idea and someone at the Department of Health said, oh yes, this is another hoop to make GPs jump through'.⁸¹ Coding a patient as having depression but then not using a standardised depression questionnaire to assess them would result in a lower QOF score and, thus, a loss of income for practices. However, the use of standardised depression questionnaires took up time in the consultation. The authors found that this set of conditions resulted in some GPs being reluctant to code people as having depression to avoid having to use a questionnaire, thereby saving time in the consultation. As one GP explained, 'I think we stop and pause a little bit before we actually put the depression code in. And, of course, there was a mad scramble around the Read codes to find a Read code that wouldn't get picked up by the QOF'.⁸¹

This study suggests that GPs did not perceive the use of standardised depression measures to be valid or necessary tools in their management of patients with depression. However, if GPs did not use them, they ran the risk of losing income. Under these circumstances, GPs avoided the potential loss of income through manipulating the ways in which they coded patients suspected of having depression. In terms of the theory under test, this study suggests that if providers perceive that the quality indicators that are subject to public reporting do not reflect what they perceive as good-quality patient care but they are financially incentivised to fulfil them, they may resort to the gaming and manipulation of data to avoid both having to fulfil them and the consequent loss of income.

Mitchell *et al.*⁸⁰

This study also explored the impact of the QOF on the diagnosis and management of depression in primary care. The authors took a purposive sample of four GP practices and conducted one focus group in each practice. Focus group participants included GPs, practice nurses, community nurses, primary care mental health workers and practice managers. They were asked to describe how the introduction of the QOF and NICE guidelines had influenced how depression was diagnosed and managed.

The authors observed that GPs found the use of the PHQ-9 (a standardised depression rating scale) time-consuming to use during the consultation and had adapted how they administered the questionnaire to fit with their consultation style. These included letting the patient self-complete the measure in the waiting room, reading out the questions to the patient and recording the answers themselves, recalling the questions from memory during the consultation and working out the score afterwards, and telephone administration of the questionnaire. The authors noted that a number of these 'workarounds' may have compromised the validity of the PHQ-9. GPs perceived that the PHQ-9 did not facilitate the clinical management of the patient, as they preferred to rely on their 'gut feeling' to determine how depressed a patient was, which they felt was often not reflected by the PHQ-9 score. Rather, the impetus to use the PHQ-9 was 'the potential for missed targets'.⁸⁰ The authors report that the financial incentives attached to the QOF acted as a 'disincentive to

code depression if a PHQ-9 was not completed by the patient'.⁸⁰ Instead, GPs used alternative codes, such as 'low mood' or 'stress', to avoid recording 'mild' symptoms as depression. As one GP explained 'diagnoses of what would be "QOF-able" depression has probably dropped . . . we realised if we kept labelling people as depressed when they perhaps weren't, then we weren't going to see them again and lose the points'.⁸⁰

This GP is referring to a scenario in which a patient with mild depression or stress consults their GP only once and does not return to the practice for follow-up, perhaps because their symptoms have resolved. In this situation, if the patient was initially coded as having depression but then did not return for a follow-up consultation, they would not be able to complete a PHQ-9 questionnaire at follow-up, the practice QOF score would go down and the practice would lose income. It reflects the clinical uncertainty regarding the diagnosis of depression and how the QOF created a penalty for getting this diagnosis 'wrong' that did not exist before its introduction. In response, GPs avoided coding patients as having depression. In terms of the theory under test, this study suggests that attaching financial incentives to quality of care indicators may create perverse incentives unless the quality indicators contribute to patient management and allow for the clinical uncertainty inherent in the practice of medicine. In situations where perverse incentives are created, this may lead to gaming or effort substitution. It also highlights the tensions between the use of PROMs as a tool to reward good practice at an aggregate level and their use as individual patient management tools.

Theory 7 summary

This collection of studies, using a range of different methods, has provided a useful test of the theory that financial incentives may amplify the impact of public reporting on QI but may also have a detrimental impact on non-incentivised aspects of care. A number of studies suggest that greater improvements in the quality of patient care occur when providers are subjected to both financial incentives and public reporting than when they are subjected to either initiative alone.²¹⁵⁻²¹⁷ Another set of studies, largely qualitative, suggest that feedback of performance indicators to providers who are subjected to neither public reporting nor financial incentives, rarely led to formal or sustained attempts to improve the quality of patient care, particularly when providers did not trust the indicators themselves.²¹⁹⁻²²¹ Under these conditions, the feedback of performance data were more likely to lead to providers improving the recording and coding of data, which may be an important first step in increasing their trust in the data itself, as well as providing a basis on which further QI initiatives may occur.

However, the evidence also suggests that financial incentives have only a short-term impact on QI if they are used to incentivise activities that providers already perform well in and when providers reach the threshold at which they would receive the maximum amount of remuneration.²¹⁸ Furthermore, there is also both quantitative²¹⁸ and qualitative evidence¹⁸ to indicate that financial incentives, together with public reporting, may lead to 'tunnel vision' or effort substitution, that is, focusing on aspects of care that are incentivised to the detriment of care that is not, especially when providers do not feel that the indicators adequately capture quality of care. There is also evidence to suggest that when providers are subjected to both public reporting and financial incentives attached to these indicators but they do not feel the indicators are valid or contribute to patient care, this can lead to the manipulation or gaming of the data.^{18,80,81} This is not always or necessarily the result of active attempts to 'cheat' the system on the part of providers. Rather, the use of financial rewards can create perverse incentives that are at odds with the inherent clinical uncertainty of conditions such as depression. Under these conditions, clinicians have to find a way to manage this clinical uncertainty at the same time as ensuring that they are not financially penalised for doing so.

Theory 8: the perceived credibility of performance data influences providers' responses to the feedback of performance data

In *Chapter 3*, we highlighted a number of theories that suggested that data must be perceived as credible and must be trusted by providers if they are going to respond to them. The previous section of this report on financial incentives revealed that, unless the recipients view performance data as valid and relevant to the clinical care of patients, financial rewards attached to their feedback can create perverse incentives to

meet targets at the expense of clinical care, and may lead to gaming and effort substitution. Benchmarking theories postulate that the lower an organisation's acceptance of poor benchmarking scores and the more data can be regarded as a 'soft indicator', the more likely it is that it will respond by denouncing the validity of the indicator and/or improve presentation of data rather than improving performance.^{93,94} If data are not perceived as valid, it is unlikely that clinicians will respond by making changes to clinical care.

One theory to explain why clinicians do not trust data lies in the methodological aspects of the indicators themselves, their coverage and validity and the process of case-mix adjustment. In the USA, a particular bone of contention is the formulation of indicators based on routinely collected administrative data gathered by insurance companies at patient discharge to bill payers, which are deemed by many clinicians to be inaccurate, versus the use of data extracted from patient notes by hospital representatives, requiring additional resources to obtain. In contrast, PROMs and patient-reported experience data are based not on clinical or administrative data but on patients' own reports of their health and experience. However, providers may question if the subjective reports of patients can serve as reliable indicators of their health outcomes, and some have expressed concerns that patients' ratings of their outcomes may be unduly influenced by their experiences.^{97,194,223} The underlying assumption of all of these claims is that it is the data and what is done to the data (i.e. case-mix adjustment) that providers object to.

Decisions about what data are collected and how they are manipulated to form indicators are made by those who design and initiate such reporting schemes. Those who initiate or mandate such public reporting of performance initiatives have a particular set of hopes and aspirations regarding the outcomes of such a scheme. An alternative, although not mutually exclusive, theory is that providers do not trust the underlying driver of the feedback and public reporting programme and question the designers' anticipated outcomes. The studies reviewed in *Chapter 4* on mechanisms revealed a tension between the idea that the goal of public reporting is to put pressure on providers to improve quality through increased competition to improve their market share versus the idea that the goal of public reporting is to improve quality through sharing data and learning from other organisations. Benchmarking theories have also drawn attention to two competing aims of benchmarking activities: competition versus collaboration.⁹¹ Wolfram-Cox *et al.*⁹¹ hypothesise that whether benchmarking is collaborative or competitive depends on structural factors such as the extent of interdependence between partners; the degree of geographical separation; the number of partners involved; and dynamic factors, for example who initiates the benchmarking, the primary motivation for initiating the benchmarking and the nature of the existing relationships between the organisations.

In the following section, we review a number of studies to try to unpick the relationships between, and relative importance of, the source of the data, the nature of the indicators and the perceived motivation behind the reporting of performance data. We start by considering studies that have attempted to understand the determinants of the success or failure of public reporting and feedback initiatives, with a particular focus on studies exploring how these factors contribute to the perceived credibility of these data.

Theory 8a: the perceived credibility of the performance data influences providers' responses to performance data

Bradley *et al.*²²⁴

This study aimed to identify successful strategies and common difficulties in implementing data feedback initiatives in hospitals. The authors focused on exploring hospitals' efforts to improve beta-blocker use after acute myocardial infarction and purposively selected eight hospitals from across the USA whose performance in this clinical area varied substantially. They conducted semistructured interviews with between four and seven staff at each hospital; in total, 45 participants were interviewed, comprising 14 medical staff, 15 nursing staff, 11 staff with responsibility for quality assurance and five senior administrative staff. Interview questions focused on how staff had collected and used the data and the degree to which they perceived the data had been effective in improving care. The authors' analysis focused on identifying 'what worked' and what did not in collecting and implementing these data.

The authors identified seven key themes underpinning what made data feedback effective; three of these themes focused on the credibility of the data. The authors report that medical staff at every hospital felt that the data must be valid and perceived as valid by clinicians in order to have any impact on doctors' behaviour. If data were perceived as valid, clinicians were less likely to reject or 'argue with' them and were more likely to respond to them. However, participants also recognised that gaining clinicians' trust in the credibility of the data took time and required effort. Strategies used to increase the credibility of the data included nurses sitting alongside doctors to demonstrate that they could accurately abstract information from patients' notes in order to create the feedback reports and to investigate any perceived inaccuracies with the data quickly 'until we're sure it's clean'.²²⁴ Finally, participants also explained that the timeliness of the data and the ways in which they were presented were also central to their perceived credibility. In particular, participants felt that collecting 'real-time' data that were 'no more than 3–6 months old' and ensuring that they were presented by someone who was 'clinically competent' were important ingredients in maintaining the credibility of the data. This study was based on the views of staff in only eight hospitals across the USA; however, it provides some initial evidence to unpick what is meant by and what supported data credibility. In terms of the theory under test, these findings suggest that the credibility of the data depends on the processes through which these data are collected and presented.

Mehrotra *et al.*¹⁹⁵

Mehrotra *et al.*¹⁹⁵ conducted interviews with 17 employers and 27 hospital managers to explore their views of and responses to employer initiated report cards in 11 regions of the USA. The authors attempted to include hospital representatives who were supportive of, as well as those who were opposed to, report cards. Hospital managers were either chief executives or QI directors. The aim of the study was to explore the determinants of successful report card efforts and to understand why some report card initiatives failed.

The authors found a mix of successes and failures in terms of whether or not report cards were perceived to have stimulated QI in hospitals. In communities where report cards had been successful, interviewees felt that there had been increased attention to quality and an increase in the presence of quality directors at board of trustee meetings. They identified a number of contextual factors or system tensions that prevented the success of report cards. They found considerable ambiguity and tension between these different stakeholders concerning the purposes of report card initiatives. In some communities, hospitals were unclear what the purpose of the report card was. In others, there was tension between employers, who were perceived to have introduced report card systems to reduce costs, and clinicians, who felt that improving quality should be the primary goal of report card initiatives.

The authors also identified conflicts regarding how quality was measured, including concerns about case-mix adjustment, whether outcomes or process measures should be used, and the validity and cost of data used to produce report cards. In terms of case-mix adjustment, hospital leaders felt that it would never be possible to develop reliable methods of case-mix adjustment, while employers felt that imperfect case-mix adjustment was better than none. Hospital leaders felt that outcome measures did not enable them to identify the source of the quality problem, while employers felt that it was a hospital's responsibility to undertake additional work to identify this. The authors found that the most 'contentious' issue between hospitals and employer coalitions was the data used to generate report cards. Many employer report card initiatives used administrative billing data that hospital leaders considered inadequate for QI, as they perceived these data to provide financial rather than quality information. Employers were more accepting of the use of administrative data. Hospitals preferred the use of clinical data to produce performance indicators but were frustrated that they had to pay for these data to be abstracted from patients' notes in order to produce report cards that they did not want in the first place. Finally, the authors found that the degree to which hospitals were involved in report card design and modification influenced their acceptance of the data.

The authors concluded that they could find no consistent set of report card characteristics that predicted which report cards were successful in initiating QI activities by hospitals, apart from the finding that successful report cards did not use administrative or billing data. They hypothesise two possible

explanations for this finding: (1) that hospitals ignored the report cards because they dismissed administrative data as inaccurate, or (2) that hospitals were not involved in the design of such cards and, as such, felt little ownership over the purpose of the scheme.

Boyce *et al.*⁵⁸

This study explored Irish surgeons' experiences of receiving peer-benchmarked feedback, replicating the same measures used as part of the UK national PROMs programme for hip surgery. However, unlike in the UK national PROMs programme, the feedback provided to surgeons in this study was at the individual surgeon level, rather than at the provider level. The feedback was not being implemented routinely and was not publicly reported, but was a 'one-off' private feedback intervention. The format of the feedback was also different from that provided by the national PROMs programme; rather than receiving a funnel plot, surgeons received a 'caterpillar plot' that graphically presented the average health gain on the OHS plus 95% confidence intervals for all surgeons (anonymised), with their own score highlighted. In this way, the surgeons were able to see how their own score compared with that of others.

The paper reports on a qualitative research study that was nestled within a larger RCT of PROMs feedback, the Patient Reported Outcomes: Feedback Interpretation and Learning (PROFILE) trial, the results of which were reviewed in the previous chapter.¹⁷⁴ This study aimed to evaluate the effectiveness of the NHS PROMs programme methodology for surgeon-level feedback in an Irish context. PROFILE tests the hypothesis that surgeons who received benchmarked PROMs feedback will have better future outcomes than those who do not. Surgeons were randomised to the intervention arm of the PROFILE trial and received peer-benchmarked feedback. All 11 surgeons in this feedback arm of the trial were invited, and agreed, to participate in face-to-face interviews. The participants varied in terms of the setting of their usual workplace, their relative performance ranking and their previous experience of using PROMs. The interviews explored surgeons' experiences of using PROMs, their attitudes to using PROMs as a peer benchmarking tool, the methodological and practical issues with collecting and using PROMs data and the impact of the feedback on their behaviour.

The authors found that surgeons had conceptual and methodological concerns about the use of PROMs data, which led the surgeons to question the validity of these data. Unlike other performance indicators, PROMs rely on the subjective judgement of patients, and surgeons questioned patients' ability to report on issues such as pain and function. Surgeons also confused PROMs with patient experience; they assumed that PROMs captured, and would be unduly influenced by, patients' experiences of their care, and they were also concerned that patients may either underestimate or overestimate preoperative and postoperative outcomes. Furthermore, many surgeons expressed disbelief about the percentage of patients who reported that they had not improved or had had problems after surgery, as these figures did not match their clinical experience and the verbal feedback received from patients. They also expressed concerns about the impact of patient case mix and differences in hospital resources and levels of community support that may affect comparisons between surgeons or providers. They also questioned the timing of PROMs follow-up and did not feel that 6 months' follow-up would fully capture the full benefit of the operation.

The surgeons also had difficulty interpreting and understanding the meaning of the data. They felt that PROMs feedback alone was not sufficient to provide an explanation for poor performance and that it did not enable them to identify opportunities for QI. This was because the surgeons perceived there to be a number of causal factors that may lead to poor PROMs scores, and thus felt that the PROMs scores did not, in themselves, highlight which of these factors required addressing. This relates to audit and feedback theories, which hypothesise that feedback must unambiguously provide information on the cause of poor performance and identify ways in which it can be rectified. The study also highlighted a number of practical issues around collecting and using PROMs data that created barriers to positive engagement with the exercise. Data collection added to workload pressures, and many surgeons stated that their supporting staff were not willing to accept this added workload. Political will at a hospital and system level were thought of as important in order to sustain any QI as they required local resource flexibility. There were also concerns about training in the use of PROMs.

The study also sought to understand how surgeons' attitudes to PROMs data related to their use of these data for QI activities. The authors' analysis identified three distinct groups of participants in terms of attitudes towards the data (typology): advocates, converts and sceptics. The advocates expressed a positive attitude towards the feedback they received, which they believed had an impact through promoting a reflective process focusing on their clinical practice. However, specific changes to care were not discussed. The converts were uncertain about the value of PROMs, and this reduced their inclination to use these data. This group generally felt that it is important to know what patients think about their outcome, but emphasised the need to provide actionable feedback. The sceptics believed that the PROMs feedback they received was not clinically useful and had no impact on their behaviour. They felt that there were too many methodological concerns to trust these data, and that these data did not provide a useful source of ideas to stimulate QI.

In terms of the theories under test, this study suggests that surgeons questioned the validity of PROMs data because they mistrusted the idea that patients' subjective experiences formed a valid indicator of the quality of care, and because they felt that the instruments themselves, the timing of measurement and the ways in which the data were adjusted for case mix did not provide an accurate indicator of the quality of patient care. This study provides some support to the theory that, owing to the multiplicity of factors that may be causally linked to an outcome, providers find it more difficult to identify the possible causes of poor outcomes.

Theory 8a summary

Two of these are small qualitative studies that rely on the self-report of hospital staff in selective regions of the USA. The other is a small qualitative study of surgeons' experiences of and attitudes towards PROMs data. However, they all suggest that both the source of performance data and the process through which the data are collected and presented are important influences on whether or not performance data are perceived as credible by clinicians. Mehrotra *et al.*'s¹⁹⁵ study in particular highlights that clinicians perceived data from patients' notes to be more credible than report cards based on administrative data. This suggests that the source of performance data is an important determinant of their perceived credibility.

Theory 8b: the source of performance data influences providers' perceptions of their credibility

We can test the theory that report cards based on data from patients notes are perceived as more credible than report cards derived from administrative databases by comparing two of the oldest cardiac reporting systems in the USA: the California Hospital Outcomes Project (CHOP) reports, which are based on administrative data, and the NYSCRS, which is derived from clinical data abstracted from patient notes. Participation in both systems is mandated by state law. Despite the age of the systems, there are a number of key differences between the two, in terms of how they were developed, the data used to produce the reports and the reporting level, that provide a useful comparison for our synthesis.

The CHOP reports, which began in 1993, are based on routinely collected administrative discharge data, and are overseen by a government agency, the Office of State Wide Health Planning and Development. The risk-adjusted data are aggregated to the hospital level only. The initial report classified hospital performance for acute myocardial infarction as 'better' or 'not better' than expected, while the second report classified hospital performance as 'better', 'worse' or 'neither better nor worse' than expected. The NYSCRS was initiated in 1989, partly in response to the shortcomings of the HCFA mortality data report cards. The NYSCRS was developed as a collaborative venture between the New York State Department of Health Authority and the 21-member-appointed Cardiac Advisory Committee. The reports were produced from chart data collected specifically by the hospital for the reports that are aggregated on a yearly and 3-yearly basis at both hospital and surgeon level. The reports contained the number of deaths, observed mortality rates, expected mortality rates and risk-adjusted ratios, and enabled the identification of hospitals and surgeons with statistically higher and lower rates than expected given their case mix.

There have been several studies exploring both systems, and we now review those that have explored clinicians' attitudes to and use of data produced by these reporting systems. It is important to note that many of these studies are based on surveys of providers, which are subject to the risk of response, recall and social desirability bias. Providers who respond to surveys may be those who have an especially positive (or negative) view of public reporting, and what providers say has happened may not always accurately reflect what has actually occurred. Nonetheless, taken together, the surveys do provide some evidence with which to test our theories.

California Health Outcomes Project reports

*Luce et al.*²²⁵

The authors surveyed 17 acute care public hospitals 1 year after the initial CHOP reports were first published to explore whether or not and how the hospitals had used the reports to initiate QI initiatives. They provide no information on the number of hospitals in their initial sampling frame, so it is difficult to determine how generalisable their findings are. The authors found that 'few, if any' QI activities were initiated in response to the CHOP data. The free-text responses to their survey suggested the main reasons for this were hospitals perceiving their outcomes to be adequate, questioning the validity of these data (as there were 'too few' patients in each diagnostic category), not having the resources (we do not know whether this refers to examining the data or addressing any issues) and not being concerned about the public release of data. In part, these findings can be explained by the fact that this study was conducted early in the history of public reporting of performance, so familiarity with these data and expectations about their use may not have created the same pressure on hospitals to respond. In line with theories discussed in the previous section on market competition, the authors also point out that public hospitals do not have to compete for patients because most patients attending are uninsured and have little choice of hospital. As such, public hospitals had less incentive to improve the quality of their care. They also explain the findings by highlighting two issues of relevance to the theories tested here: (1) that providers struggled to understand these data and (2) that they distrusted the method used to risk adjust these data. In their conclusions, Luce *et al.*²²⁵ noted that, at the time of writing, 'hospitals continue to resent the fact that they are required at their own expense to provide the Office of State Wide Health Planning and Development with discharge data that can be used against them in the competitive medical marketplace'.²²⁵ Here we see that providers distrusted not only these data themselves, but also the perceived motivation behind the data's collection: creating 'winners' and 'losers' in a competitive market.

*Rainwater et al.*²²⁶

This survey was conducted 2 years later than that carried out by Luce *et al.*²²⁵ The authors surveyed 249 hospitals (out of the 374 who received the CHOP report) and then interviewed a purposively selected subsample of 39 hospital quality managers from the state to explore how they had used the second publication of the CHOP reports for QI purposes. They found that managers expressed concerns about the quality of data coding on which the reports were produced and whether or not the report provided a valid comparison of dissimilar hospitals. When respondents were asked what they found least useful about the report, the most frequent response was that the report 'was not timely and did not reflect current practices'.²²⁶ The respondents also felt that the report provided information outcomes but not 'practical information about the process of care',²²⁶ which they regarded as key information for driving QI. The QI managers wanted to know what better-performing hospitals were doing differently. The respondents also indicated that quality information they obtained from other sources was more useful than the CHOP reports and cited systems that were characterised by process data and rapid feedback. Some felt that the CHOP reports simply confirmed what they already knew from other data.

Similar to Luce *et al.*,²²⁵ they observed that two-thirds of respondents had taken no specific action in response to the reports, although the reports were disseminated widely among hospital staff. Of those who had taken action, responses included (1) review of care and instigation of new care pathways, (2) changing medical staff and (3) improving process of data coding. Interview participants explained that CHOP data had been useful for improving hospital coding and for educating doctors about the importance

of coding, because this affects the compilation of the indicators. The authors conclude that the public reporting of performance 'although not completely ignored, is not a strong impetus for change or improvement in the process of care'.²²⁶ The authors observed that hospitals typically responded in a way that lies 'between these two extremes and can be viewed as largely ceremonial. Organisations responding in a ceremonial manner alter observable activities to create the impression that established processes are working, without actually altering core activities',²²⁶ namely patient care. This conclusion resonates with van Helden and Tilemma's⁹³ benchmarking theory, that when organisations do not accept the validity of the indicator, they are more likely to respond by improving the presentation or appearance of these data rather than improving performance. However, these findings can also be interpreted as indicating that providers' initial responses to the report cards focused on efforts to improve the validity and credibility of these data, through improving the process of hospital coding.

New York State Cardiac Reporting System

We now turn to the NYSCRS. These report cards are based on clinical data abstracted from patients' notes and were overseen by a committee of cardiologists and cardiac surgeons. On this basis, therefore, we might expect that the cards would have been better received than the CHOP reports. One survey directly compared provider responses to the CHOP and the NYSCRS reporting systems, to test the theory that reporting systems based on clinical data were viewed more favourably by providers than those based on administrative data.

*Romano et al.*²²⁷

Romano *et al.*²²⁷ surveyed 249 of 374 hospitals in California and 25 of 31 hospitals in New York to compare the views of hospital leaders on the two reports. Some caution is required in interpreting the findings of this survey, as hospitals with high volumes of acute myocardial infarction were more likely to respond than those with low acute myocardial infarction volumes. The authors also noted their suspicion that hospital leaders with strongly negative or positive views were more likely to respond to the survey than those with neutral views, which may have skewed the findings to the extremes. Nonetheless, the study does provide a useful comparison between the two reporting systems.

This study found that 68% of hospital leaders in California, compared with 89% of leaders in New York, agreed that risk-adjusted mortality data were useful in improving the quality of care. The New York report was rated significantly better than the California report in its usefulness in improving the quality of care, accuracy in describing hospital performance and ease of interpretation. In California, 50% of respondents agreed that the state's reporting system was better than other systems that used administrative data, while 81% of respondents in New York agreed with this statement. Conversely, 24% of hospital leaders in California agreed that their state's reporting system was better than other systems based on clinical data abstracted from patients' notes, whereas in New York 50% of respondents agreed with this statement.

Hospital leaders in New York were, in general, more knowledgeable than those in California about the methods of risk adjustment for their reporting system. However, only 8% of leaders in California and 22% in New York rated the report as 'very good' or 'excellent' in facilitating QI. This indicates that, although report cards based on clinical data may be better received than those based on billing data, hospital leaders are yet to be convinced of their value in QI. This suggests that it is not merely the nature of these data that determines their use and value in initiating QI initiatives. Indeed, the authors conclude that NYSCRS higher ratings 'may not be attributable to its use of detailed clinical data. Those ratings may, instead reflect New York's longer track record . . . [and] greater oversight by a Cardiac Advisory Committee, and a limited population of hospitals'.²²⁷ Chassin *et al.*¹⁹¹ (discussed previously under theory 3) also concluded that the durability of the NYSCRS was attributable to 'its integration into the routine process of a governmental agency . . . and the vigorous involvement of the state's leading cardiac surgeons and cardiologists in the advisory committee process'.

Theory 8b summary

These studies suggest that the CHOP reports were widely disseminated within hospitals but stimulated few, if any, attempts by providers to initiate QI activities. Instead, providers responded by taking steps to improve the validity of these data, as, by and large, they did not perceive them to be credible. This lack of credibility stemmed from the reports' reliance on administrative data, their lack of timeliness and their failure to provide information on the process of care that underpinned the outcomes data. The NYSCRS reporting system seems to have had a somewhat greater impact, with poorly performing hospitals taking steps to improve patient care. When compared head to head, the NYSCRS was, in general, better received by hospital leaders than the CHOP reports. However, the small relative advantage of the NYSCRS cannot be attributed simply to its use of clinical data, and the relative disadvantage of the CHOP reports were not solely due to the use of administrative data. Instead, studies point to the idea that clinicians distrusted the underlying rationale for collecting the CHOP data, while the NYSCRS was better received due to the involvement of leading clinicians in its design, through an advisory committee. In terms of the theory under test, this suggests that clinical involvement in the design of the report cards in addition to the nature of these data is a better explanation of their success or otherwise than the nature of these data alone. As we initially highlighted, these two conditions are not mutually exclusive and we need to understand what it is about clinician involvement in the design of report cards that influences their success.

Theory 8c: the perceived underlying driver of public reporting systems influence providers' responses

We can test the theory that the perceived underlying driver of public reporting systems influences providers' responses by comparing providers' responses with mandatory and voluntary reporting systems. Mandatory reporting systems can be initiated by regulators, national or state governments, insurance companies or employers. They may have a range of different drivers for initiating such schemes, and evidence¹⁹⁵ reviewed previously suggests that these may be at odds with what clinicians perceive to be important. In contrast, voluntary reporting systems are often initiated by professional groups or independent QI organisations whose values may better reflect what clinicians perceive to be important. We begin by reviewing studies that provide some data on hospital leaders' views and GPs' views of mandatory public reporting of performance initiatives. Some studies also report providers' views of how they have responded to voluntary versus mandatory reporting systems and their views on externally produced, publicly reported systems versus internally collected data.

Hafner *et al.*⁷⁵

This study explored provider views of the nationally standardised acute myocardial infarction, heart failure and pneumonia performance measures produced and reported by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) in the USA. The Joint Commission is a non-profit organisation that accredits hospitals in the USA, and public reporting of quality indicators forms a mandatory part of that accreditation. The JCAHO also carries out inspections to check that hospitals are meeting the minimum standards. Thirty-six hospitals were randomly selected from a sampling frame of 555 Joint Commission-accredited hospitals and invited to take part in the study. Twenty-nine hospitals agreed to participate; nine had performance indicators consistently above the mean and were 'high' performers, seven were equal to or below the mean (the authors classed them as poor performers, but it could be argued that there is a difference between being at and being below the mean) and 13 hospitals had both high and low scores ('mixed performers').

Data were collected as 29 focus groups in each hospital with a total of 201 participants, including managers and frontline staff such as doctors, nurses and administrators. The focus groups consisted of mixed groups of staff and were conducted by Joint Commission staff, who were responsible for both producing the performance indicators and accrediting the hospitals. The questions asked tended to focus on the positive elements of public reporting rather than the negative aspects. Thus, the study was at risk of numerous potential sources of reporting bias; frontline staff may not have felt able to express contrasting views to those of senior managers. The authors report that 'in interviews involving both leadership and frontline staff, more detailed responses to questions were proffered by those in leadership roles with front

line staff affirming response either with non-verbal cues or simple one word responses',⁷⁵ suggesting that 'chatty bias' might have been at play here. As such, findings are more likely to represent the views of managers than frontline staff. Furthermore, participants may have been reluctant to express negative views about performance indicators to the organisation responsible for accrediting their hospital and producing the indicators.

The paper reports largely positive impacts of performance reporting on QI initiatives. The authors found that the public reporting of performance resulted in managers, clinicians and administrators becoming more engaged in QI activities. A nurse commented that it provided justification for securing additional resources for QI activities. It also served to prioritise and focus attention on issues raised by the performance data. For high-performing hospitals, this arose from their desire to maintain this status. For low performers, the data led to an awareness of the need to critically analyse the data and respond to the findings.

Throughout the paper there is evidence that the drive to improve did not come simply from the intrinsic motivation of the staff but from the fact that the data have been made public. Participants reported that making the data public had 'drawn their attention to it' and 'forced them' to look at the data and respond to them. For example, an administrator indicated 'When you tell a surgeon his numbers are going to be out there, you get their attention and they ask what they need to do!'.⁷⁵ Similarly, nurses from a low-performing hospital indicated that public reporting had 'forced us to look at it [the data], to compare it. Before, it just sat there, now it drives us to do better' and that 'knowing that the public is aware of the scores makes us more energised to do better'.⁷⁵ The findings suggest that media scrutiny of the data put pressure on staff to respond to them, as one administrator noted: 'I don't think anyone believed it was going to be public until the newspaper article – that's when people gasped!'.⁷⁵ Media scrutiny was perhaps felt more strongly for lower-performing hospitals, as interviewees felt that the local media typically tended to focus on 'why the numbers are low and rarely on why they are high'.⁷⁵ These findings resonate with public disclosure theory, discussed in the previous chapter, which hypothesises that the media act to reinforce the shaming mechanism of public reporting and prompt the desired response.

A further issue to emerge from this paper was that interviewees in 17 of the 29 organisations reported that the validity and reliability of performance data were challenged by frontline staff. This occurred in both high-performing and low-performing organisations, but occurred more frequently in the low-performing hospitals. Concern was expressed that the data did not fully capture the quality of care in the hospital, that comparisons between hospitals did not adequately reflect differences in case mix and that the data were too old to reflect current practice. The authors noted that high-performing hospitals saw these challenges as 'learning opportunities', although they do not fully explain what is meant by this. These findings support audit and feedback and benchmarking theories, namely that when performance feedback is inconsistent with a provider's own estimation of their performance it may not be accepted.

The methodological limitations relating to the numerous sources of reporting bias may explain why this study tended to find more positive views of the impact of mandatory performance reporting than other studies. In terms of the theories under test, this study suggests that mandatory public reporting had served to place QI at the top of the agenda for providers and that increased resources had been directed to addressing quality issues. It suggests that providers were sceptical about the quality and validity of the data. It also indicates that mandatory public reporting 'forced' and 'energised' providers to address quality issues because of the media scrutiny they expected to receive or actually received.

Asprey *et al.*²²³

This study aimed to explore primary care providers' views of and responses to feedback from the national GP patient survey in England. This is a national survey of a random sample of patients registered at all GP practices in England, which asks about their experiences of care provided by their practice. The survey includes items on the ease of getting through on the telephone, the helpfulness of receptionists and being able to see your preferred doctor, as well as ratings of how good patients felt their doctor and nurse were on a number of dimensions. The survey is run by the CQC and is, essentially, mandatory. The results from

the survey are publicly reported on a special GP patient survey website hosted by the CQC, where patients can compare the results of their own practice with those of two others of their choice. In addition, the results of some elements of the survey are used in calculating GP practice QOF payments. As such, the GP patient survey is mandatory and publicly reported, and has financial rewards and sanctions attached to it.

The authors selected four PCTs to represent different geographical regions across England, and within these four areas they selected the five highest-scoring and five lowest-scoring practices on the GP patient survey item 'ease of obtaining an appointment with a doctor',²²³ an item that contributed to the practices' QOF scores. Their aim was to recruit one high- and one low-scoring practice from each of the four areas; the first practice to agree to take part within each sampling strata was included in the study. Ten GP practices were recruited, four with high scores and six with low scores, which included two single-handed practices. In each practice, two GPs (except in the single-handed practices), one practice manager and one practice nurse were interviewed, giving a total of 37 interviews.

The authors found that participants were sceptical about the credibility of the survey findings for a number of reasons. They were concerned that respondents were unlikely to be representative of the practice population because the 'vocal minority' with negative views were more likely to reply, while other groups, such as older people, working people and people with mental health problems, were less likely to reply. They also felt that the items in the questionnaire did not necessarily represent what constituted 'good' care, as not all patients valued them in the same way, or it was unclear that a high score on an item reflected a good experience; for example, a high score on waiting times might indicate that patients were being rushed through appointments in order to shorten waiting times for patients. Participants also felt that the lack of adjustment for case mix also reduced the validity, and thus the utility, of the feedback for improving patient care.

The authors report that the 'most emotive' responses provided by interview participants related to their suspicion that the surveys were driven by 'political motives'. One GP from a low-scoring practice described the ways in which the questions were asked as 'stacked against us and I think most GPs have a cynical view about that'.²²³ A practice nurse from a low-scoring practice also felt that 'this is a cost cutting exercise and little to do with a real commitment to patient satisfaction or to help those in primary care deliver a better service'.²²³ In other words, low-scoring practices in particular felt that the survey was a politically motivated attempt to cut costs rather than improve patient care. Such feelings sometimes led participants to reject and ignore the survey results, as one GP from a low-scoring practice explained:

I'm totally cynical about the government's motivation and this is just part of that . . . So if they think they've got me over a barrel, forget it, because they haven't. And I can just happily carry on and ignore this survey.

GP from a low-scoring practice²²³

One GP from a low-scoring practice commented that although the financial element of the QOF was important – 'we like to maintain a high income, of course we do' – financial concerns were not 'paramount',²²³ rather, the practice was more concerned about the 'shame factor . . . information is shared so much, you don't want to see yourself . . . on a bar chart at the bottom of the pile'.²²³

The suspicion that political motives were behind the survey was not limited to low-scoring practices; those in high-scoring practices also felt that survey was being conducted for political ends to make GPs work harder but with few clinical gains. One high-scoring GP commented that the GP patient survey was:

. . . a way of softening up primary care for extended hours, by showing there was a demand out there for it . . . I don't think there's going to be huge clinical gains from doing that.²²³

Furthermore, the questions in the survey were also perceived to measure what mattered to politicians, to the exclusion of other important aspects of care. As one GP from a low-scoring practice expressed:

*It's a bit of tail and dog isn't it? . . . Because it has been measured, is it necessarily important? It's important but is it as important as some of the things that haven't been measured?*²²³

Some participants also felt that the items in the questionnaire driven by government rhetoric, such as rapid access, had unrealistically raised patient expectations about what they could expect from general practice, but, at the same time, the government had not provided additional funding to enable practices to deliver those promises. Finally, some practices had acted on the survey findings and made changes, such as extending their opening hours, but these had not been reflected in their scores, which led to practices feeling discouraged and thinking:

Well why? What else can we do?

*Practice manager from a low-scoring practice*²²³

In terms of the theory under test, these findings suggest that this government-initiated, mandatory public reporting programme was perceived to be driven by political motives rather than a desire to improve patient care. Consequently, primary care providers perceived that the items in the programme reflected government definitions of what constituted 'good' patient care and measured what mattered to the government, rather than to primary care practices themselves. In turn, practices were sceptical about the credibility of the data as an aid to improving patient care. This feeling was further reinforced when, even after attempts to respond to the survey by making changes to patient care, these improvements were not reflected in the scores achieved on subsequent surveys, further challenging the credibility of the survey itself.

Pham *et al.*²²⁸

This paper explores provider responses to a range of performance reporting initiatives and pays particular attention to two initiatives. The first is the reporting required by JCAHO, an organisation responsible for the accreditation and regulation of hospitals in the USA; public reporting is a condition of accreditation and thus is mandatory. They also explored the Hospital Quality Initiative, which is a voluntary system launched by the CMS and is reported on the Hospital Compare website. Following the reporting system's introduction, participation in it was poor until a state law was passed that non-participating hospitals would not receive a 0.4% annual payment update. The authors also examined providers' views on other forms of reporting systems; as such, the paper provides evidence on providers to the different drivers behind a range of reporting systems, including a contrast between mandatory, voluntary and clinician-led public reporting systems.

The paper drew on data collected as part of the Community Tracking Study. The Community Tracking Study is a longitudinal study based on site visits and surveys of health-care purchasers, insurers and providers, and is focused on tracking changes in the accessibility, cost and quality of health care. The data selected for the study were collected in 2004–5 and consisted of 111 interviews with five hospital association leaders, representatives from the JCAHO and CMS and six state reporting programmes, 21 chief executive officers, 21 vice presidents of nursing, 30 quality officers and 26 clinical directors from 2–4 of the largest hospitals or hospital systems in 12 health-care markets across the USA. The data were collected using semistructured interviews that explored specific reporting programmes and their perceived impact on the hospital's organisational culture around QI, priorities, budget, data collection and review activities, feedback and accountability mechanisms. Clinical directors were also asked about their use of 11 QI tools targeted at chronic heart failure, as both the CMS and the JCAHO reporting systems included these.

The authors found that respondents mentioned involvement in 38 different reporting programmes, with each hospital participating in a mean of 3.3 programmes. These programmes varied along a number of

axes, which correspond closely to the contextual factors identified within the public disclosure, audit and feedback and benchmarking theories. These included:

- ‘Sponsorship’, or who initiated the reporting system: by purchaser, regulator, private insurer, professional groups or other private organisations.
- Data type: hospitals submit primary data for public reporting (JCAHO), primary data for private benchmarking (e.g. hospital consortia) or secondary data (e.g. insurance claims or patient surveys).
- Mandatory versus voluntary: although most programmes are voluntary, this is influenced by the nature of incentives attached to the programme – incentives explicitly tied to participation in the programme perceived as rewards (e.g. pay for performance) or punitive (e.g. loss of accreditation).
- QI support: whether or not programmes provide prescriptive information to guide hospitals’ QI activities.
- Inclusion of clinical outcome measures.

Senior hospital leaders perceived that key drivers, such as linkages to payment, JCAHO accreditation and peer pressure from public benchmarking, had raised the priority given to quality measurement and improvement by hospital leaders. This was manifest in several ways, including the inclusion of QI priorities in strategic planning, boards and senior management taking more responsibility for the formal review of performance data and associated improvement strategies and performance-related pay for chief executives. Respondents also believed that CMS and JCAHO reporting programmes had also positively changed doctors’ attitudes towards quality measurement and improvement. The accreditation and financial consequences of reporting programmes could be used as leverage by quality officers in their dealings with doctors. However, the respondents also felt that these programmes artificially focused on a limited number of indicators, which has directed both attention and resources away from other important clinical areas.

As a result of their mandatory requirements, participants felt that the drive to participate in CMS and JCAHO reporting was of a ‘push’ nature. Hospitals did direct resources at the reported clinical areas but ‘without taking standardised approaches to improving performance’.²²⁸ Motivation to be involved in other reporting initiatives was described more in terms of a ‘pull’ because they offered support by specifying changes in care processes. These forms of reporting were seen as attractive because they ‘don’t leave hospitals flailing about trying to identify evidence-based interventions on their own’²²⁸ and because they encouraged a culture of continuous QI. These sorts of programmes were most likely to be those initiated by QI organisations, state QI organisations and professional organisations.

Hospital respondents were divided on whether or not reporting had a significant impact on specific process changes to improve the quality of care. Those who had participated in QI programmes prior to being involved in mandatory public report programmes felt that public reporting had little impact on their QI interventions, as these were largely driven internally. The respondents were also divided regarding whether or not reporting had a ‘spill over’ effect on improving quality for the non-reporting conditions; some believed that QI was limited to the targeted conditions, while others felt that it had raised frontline staff’s ability and eagerness to identify and address problems in non-reported areas too.

In their conclusions, the authors argued that national programmes that mandate participation through regulatory or financial reward mechanisms ‘can influence nearly all hospitals and garner attention from those that would otherwise not prioritise QI highly’,²²⁸ whereas voluntary programmes, especially those which also provide ongoing support to implement QI initiatives, ‘help focus priorities at hospitals that are eager to take on the more challenging goals of QI’.²²⁸ Thus, national programmes with mandatory reporting and regulatory or payment consequences have increased hospital leaders’ and frontline staff’s attention to quality and directed resources towards QI. However, at the same time, such public reporting initiatives may ‘artificially narrow the scope of QI in which hospitals might otherwise engage, especially for those with long institutional histories of QI’.²²⁸ In terms of the theories under test, these findings suggest that, for those not already involved in QI activities, mandatory public reporting services serves to raise awareness of quality of care issues and direct resources towards issues raised by such reporting systems. However, similar to the findings of Mannion *et al.*,¹⁸ these can lead to tunnel vision, where other important clinical areas do not

receive the same attention. For those hospitals already involved in QI activities, mandatory reporting systems were perceived to have little additional impact on these activities. Furthermore, hospitals were attracted to systems initiated or run by QI or professional organisations because they offered support to providers to take the further steps necessary to identify the source of the problem and implement QI activities.

Davies⁹⁸

As described in *Chapter 4*, this study examined the responses of US providers based in California to both externally produced, publicly reported data and internally produced, privately fed back data, with a specific focus on cardiology. They explored providers' views of a publicly reported data system, such as the CHOP, which publicly reports data on 30-day mortality for acute myocardial infarctions and also confidential data systems designed for internal use, for example the national register for myocardial infarction by Genetech and Health Care Financing Authority, sponsored peer-review organisations in the state of California. As such, the study provides a useful contrast between provider responses to data that is produced and fed back in different ways.

This was a multiple case study of six hospitals purposively selected because they were 'high performers', as the author expected that he would be more likely to find examples of QI in those hospitals (with the assumption that the hospitals had become 'high performers' as a result of QI activities). However, a range of different hospitals was in the sample, including two academic medical centres, one health maintenance organisation, two private but not-for-profit medical centres and one public provider 'safety net' hospital. The author conducted 35 interviews with 31 individuals lasting between 54 and 90 minutes. Interviews were conducted with key informants in each setting, including the chief executive, senior clinicians with management responsibilities, senior quality managers, the chief of cardiology, a senior nurse manager, and two or three frontline staff within cardiology. As such, this study provides insight into the views of frontline as well as managerial staff.

Participants questioned the validity and reliability of publicly reported data because they perceived that the reporting systems did not take case mix into account. Participants also expressed concerns about inconsistent coding practices and the poor quality of administrative data; differences in performance were seen as artefacts of the data collection process rather than reflecting real differences in performance. This led to efforts to reform the data collection process. Participants also felt that issues that were measured attracted more attention than was warranted, to the detriment of other, unmeasured services. Quotations from respondents suggested that a reason this was perceived negatively by some staff might have been because it drove efforts to address the problem that were not necessarily clinically appropriate: 'It really fires people up to meet the task, rather than for clinically appropriate reasons'. Participants also reported instances in which clinicians challenged the pressure exerted on them by a purchasing group following the feedback of performance data, because they felt that the priorities being imposed on them by the purchasing group were counterproductive.

The author's analysis of his data suggested that a provider's action in response to publicly reported comparative performance data was most likely when external data indicated they were performing poorly: 'being an outlier does motivate performance. There's no doubt about that'.⁹⁸ A response to these data was less likely if they were a 'middle ranker' because some providers were willing to tolerate being 'middle of the pack' and did not feel that they had an incentive to improve. However, in some instances being 'middle of the rank' was not acceptable when providers were 'striving to be the best', while, for others, a provider taking action depended on 'our own perception as to whether [the data] were an accurate reflection of what we think is happening'.⁹⁸ The paper also noted that publicly reported data were seen as only one source of information about the quality of care, with their own assessments and the views of peers and coworkers being as important. For example, one senior clinician noted, 'It's the opinion of peers that matter more than anything else about quality' and another senior clinician explained that comparative performance data 'merely reinforces already held opinions just based on other factors, you know, day to day experience'.⁹⁸

The study also sheds light on the relative roles of externally produced and internally collected data on the implementation of QI initiatives in response to these data. The findings suggest that publicly reported

data focused attention and acted as a 'kick start' to the QI process; as a senior nurse manager explained, 'external data are the start of the process . . . that really gets the ball rolling in terms of an [internal CQI] investigation'.⁹⁸ However, external data were not able to identify the cause of the problem and thus could not help to identify a solution because these data were not timely and thus lacked relevance to current care – 'If you're not doing it yourself and reacting to it immediately, there's a whole time lag and opportunities for improvement that you've missed' – and did not provide sufficient detail: 'you just don't get the details [from the external data]'.⁹⁸ It was the internally collected data that provided the necessary clinical detail to identify the cause of the problem and how to remedy it: 'it's the in-house data [that] drives us more than the outside data. I think it's also better data and it's more focused; it has many more elements to it' and 'our best successes [in using data to improve quality] were our very own internal ones'.⁹⁸

Therefore, external data were useful to identify *what* needed to be looked at, but internal, clinically owned process-based data were needed to identify the cause of the problem and how it could be dealt with. To support this, providers also needed practical resources for the analysis, presentation and interpretation of such data and a culture that valued and supported continuous QI processes: 'we have wonderful motivated people but if we didn't have the resources to do this, we couldn't. There's not only people committed to excellence, there's resources committed to excellence'.⁹⁸ If good local data and supportive resources were absent, little QI was seen: 'We don't do it [benchmarking] and we don't have the resources to do it . . . really, no way, since we don't have ongoing databases'.⁹⁸

Thus, in terms of the theories under test, this paper suggests that when publicly reported data highlighted issues that the clinicians themselves or their peers also felt were a problem, they served to amplify or kickstart an intrinsic desire to improve. It also indicates that although publicly reported data might focus attention on areas that need to be changed, it was only through analysing internally collected process data that providers could understand the cause of the problem and identify ways to address it. To be able to do this, providers needed practical resources and management support for internal data collection and the analysis and interpretation of those data.

Theory 8c summary

These studies suggest that mandatory public reporting systems had focused the attention of hospital leaders and frontline staff on quality issues, particularly for those who had no previous experience of QI activities. However, unless the issues raised were also identified as a problem based on clinicians' day-to-day experience, by their peers or by internal data collected by the hospital, the focus on the indicators included in mandatory public reporting systems was perceived as leading to 'tunnel vision'. In particular, for those who were already engaged in QI activities, public reporting could artificially focus attention on a limited number of issues, at the expense of other clinically important areas. Furthermore, although scoring poorly on an indicator included in a public reporting system could 'kick start' a response from providers, it was only through analysing internally collected data that providers could understand the source of the problem and identify a possible solution to rectify it. However, providers also reported that considerable resources were required to enable them to do this. It is likely that those who were already engaged in QI activities had set up internal data processes that inform QI on an ongoing basis and thus were better able to respond when external data shone a light on poor performance. It is therefore unsurprising that providers valued the additional support for QI activities that was offered by reporting systems led by clinical or QI organisations and felt 'pulled' rather than 'pushed' to engage with these reporting systems.

Theory 8d: clinicians have greater trust in clinician-led reporting systems

To test the theory that clinicians have greater trust in clinician-led public reporting systems, we now look in more detail at one programme that was initiated and led by a collaboration between clinicians, hospital providers, insurers and employers: the WCHQ.

Greer¹⁹⁹

All quotations from this study are reproduced from Greer,¹⁹⁹ with permission from the Commonwealth Fund.

Greer's report of how the WCHQ was set up is overwhelmingly positive about its impact; this may be due, in part, to its reliance on data collected from the 'enthusiasts' who led the collaborative. The author recognises that, as the collaborative was a 'pioneering effort', it benefited from the Hawthorne effect. The report is based on 31 interviews with board members and medical practice executives involved in the WCHQ. They included eight CEOs, 10 chief medical officers and five executives responsible for the quality of care in their organisation. As such, it does not capture the perspectives of staff on the ground or their responses to this initiative. Nonetheless, the study provides some useful ideas about the mechanisms through which clinical engagement and trust in a public reporting system was achieved.

As Greer describes, the WCHQ was founded in 2003 by chief executives of several large multispecialty practices and their partner hospitals. The collaborative was brought together with the purpose of promoting QI among member organisations by (1) developing performance indicators, (2) openly sharing provider performance through public reporting and (3) identifying and sharing best practices to improve the performance of all members of the collaborative. One of Greer's interviewees, a doctor by background who was the CEO of a doctor-led health-care organisation, noted:

What we sensed was that there were reports being published by people who had some knowledge, but perhaps not full knowledge of healthcare and its delivery . . . The reason for our meeting was: it looks like people are going to start writing reports, publishing reports on medical performance and that will be followed by dictating type of care and how care should be delivered. Shouldn't we, the people responsible for care delivery, shouldn't we be involved in the process?

p. 2¹⁹⁹

In other words, the WCHQ was set up in order to allow clinicians to retain some control over the ways in which public reporting was designed and implemented. The underlying premise of this process was perceived to be one of mutual learning, as an antidote to the competitive climate providers perceived they were expected to work in. One of Greer's respondents, a founding member of the WCHQ, noted:

*I think in many ways WCHQ became an oasis from this highly competitive environment . . . a safe harbour where we are not talking about market dominance and control. We are talking about quality.*¹⁹⁹

A chief medical officer who joined the initiative later, when the opportunity for membership was opened up to a wider group of clinic groups and practices, also remarked:

*[I enjoy] the sense of collaboration, and what is kind of fascinating, is that the discussion – of how we are doing, how we are doing relative to each other, how we can do better – constantly brings you back to your primary purpose and that is the patient you are taking care of.*¹⁹⁹

p. 13¹⁹⁹

Here we see that the primary motivation for signing up to the WCHQ was its focus on how to improve patient care, rather than on improving profits through market dominance.

The mechanisms of change reported by interviewees were that (1) doctors were intrinsically motivated to do a good job, and receiving feedback that indicated their performance was poor prompted them to take steps to improve; and (2) doctors can identify what needs to be improved, and how, by sharing and learning from the best practices of other providers. These mechanisms resonate with both audit and feedback and collaborative benchmarking theories. Greer's interviewees noted:

I think one of the constructs that the Collaborative built on is that physicians want to do a good job . . . By providing information about how physicians perform you can influence physicians behaviour . . . They are driven to change things when their performance does not look good.

p. 14¹⁹⁹

WCHQ [provides] the actual benchmarking data for looking at where you are at, how to improve, and building on those connections with other organisations that are similar to you; where you can say 'our numbers are not good here, how did yours get better? What can we learn from how you are doing it?

p. 17¹⁹⁹

Greer argued that the WCHQ had a high level of clinical engagement, as evidenced by its growing numbers, with 50% of Wisconsin primary care physicians being members of the collaborative in 2008. Greer attributed this high degree of clinical engagement to the fact that the collaborative was clinician led, rather than led by government or an insurance company. This resulted in the development of indicators and methods of attributing those indicators that were perceived as valid reflections of clinical care. The indicators were accepted because there was a shared perception among the collaborative that the motivation behind the production of the indicators was that of improving patient care, rather for competition or external regulation. Participants in the WCHQ indicated that they accepted the data collated by the WCHQ rather than dismissed them because they were developed by clinicians and as such were seen as an accurate reflection of clinical care. The fact that the collaborative developed the indicators made it much more difficult for its members to then dismiss their validity. As one of Greer's interviewees reports:

... we promised each other we would report our data, we would not fudge it, we would have it verified, we would make it public and we would not walk away from whatever we found.¹⁹⁹

In contrast, Greer's interviewees were sceptical of the motivation behind insurer, claims-based indicators that were perceived as inaccurate and designed for competitive advantage rather than claims purposes. Indicators developed by organisations that were representatives of the government, such as the Quality Improvement Organisation or the JCAHO, were seen as motivated by external regulation and were perceived to have only a short-term impact on QI initiatives, as one of Greer's interviewees observed:

JCAHO is big brother coming in and the response is usually: 'Oh there is a JCAHO visit. We will get ready for the JCAHO visit that we will have to pass.' Then they leave and they don't come back for 5 years and until the next one comes along nobody thinks about them.

p. 17¹⁹⁹

In summary, Greer's case study of the set-up of the WCHQ suggests that clinicians were engaged in setting up the collaborative because they supported its primary motive of improving the quality of care by comparing performance and sharing best practices. The perceived mechanisms of change were those evoked by audit and feedback and benchmarking theories; that is, clinicians have an intrinsic motivation to do a good job and will be prompted to improve if their performance is poorer than they would like or poorer than that of their peers. This engagement, in turn, led to the development of indicators that reflected the information that clinicians needed to improve patient care. As clinicians were involved in the design of the indicators, it was then difficult for them to dismiss the data as invalid. This suggests that clinicians are more likely to trust data that are publicly reported if they have a role in the choosing the indicators and the means through which they are risk adjusted and reported.

Lamb et al.²²⁹

This paper reports on a longitudinal cohort study and survey of members of the collaborative to explore whether or not the WCHQ led to improvements in the quality of care provided. For each publicly reported indicator, the authors examined whether or not there was an improvement in mean performance for the collaborative as a whole over time and examined trends over time. Clinician groups were ordered by their rank in first year of reporting and this was correlated with their rate of improvement. They conducted a postal survey of clinic groups to examine whether or not QI projects were undertaken specifically in response to reporting. Finally, for four indicators for which comparative data were available, they compared improvement in care for patients in the WCHQ with (1) patients in Wisconsin who were not part of the collaborative, (2) patients in Iowa and South Dakota, where there was no public reporting, and (3) residents

in the remainder of the USA. Their analysis is based on responses from 17 out of the 20 doctor groups who were part of the collaborative, representing 409 out of the 582 clinics.

The authors found that for the WCHQ as a whole, for each reported measure, the indicators for clinics that were part of the collaborative improved as a whole over time, but there was wide variation in the amount of improvement, from 1.2 for low-density lipoprotein control to 17.3 for monitoring kidney function. Not all of these improvements were statistically significant (of 13 measures, 7 showed a statistically significant improvement over time). Groups that were initially low performers improved at a greater rate than those that were initially high performers. The survey found that 15 out of 16 groups reported formally giving priority to at least one QI measure in response to reporting. Nine out of 16 groups indicated they always or nearly always set priorities in response to reporting, while 6 out of 16 indicated that they sometimes did so. The mean number of QI interventions initiated by members of the collaborative increased over time. On three of the four measures for which comparative data were available, there was a trend for patients in the WCHQ to receive better care than patients in Wisconsin who were not part of the collaborative, in Iowa and Dakota and the USA as a whole, but this was not statistically significant. The measure that WCHQ members did not perform better on was not publicly reported by the WCHQ.

It is not possible to attribute the improvements found in this study solely to public reporting, as improvements in the quality of care were also found in other locations where public reporting did not occur. As the authors acknowledge, clinicians in the WCHQ were volunteers and more likely to be enthusiastic about public reporting and their patients were more affluent. Nonetheless, these findings suggest that members of the WCHQ did take steps to improve patient care in response to public reporting. They also indicate that providers' responses were variable, suggesting that no group was able to respond to all the reported measures and clinician groups prioritised a limited number of indicators to focus on. Of note here is the finding that those initially identified as low performers improved at a faster rate than average or high performers. The authors hypothesise that 'public reporting creates a milieu in which parties compete for external recognition and strive to avoid the negative aspect of publicly being identified at the bottom of the list'.

Smith *et al.*²³⁰

This paper reports on survey of doctor groups who were members of the WCHQ to explore whether or not they had initiated any improvements in diabetes care in response to the collaborative's public reporting. The authors invited 21 doctor groups, representing 582 clinics, to participate, of which 17 groups representing 409 clinics agreed to participate. They received group responses from 231 clinics and individual responses from 178 clinics. They carried out two surveys: one for the doctor group as a whole and one for each clinic. They asked the clinics if they had implemented any of 55 QI initiatives, of which 22 were diabetes related, in each year between 2003 and 2008. The groups were asked to identify any year in which a metric was chosen to be a focus for QI and to indicate if this was in response to the public reporting. It is unclear how this was quantified, for example whether it was a simple yes/no answer. They used these data to generate an indication of whether, for each year, the doctor group formally adopted a focus on one or more diabetes metrics in response to the collaborative's reporting, whether they adopted a focus but not in response to public reporting or whether they did not adopt a focus. Given that this survey is based on self-report, there is a risk of reporting bias in terms of both social desirability and recall bias.

They found that the implementation of diabetes QI initiatives increased between 2003 and 2008. Clinics in groups that focused on diabetes metrics in response to public reporting were more likely to implement both single and multiple initiatives than groups that did not formally adopt a focus on diabetes. In this group, a factor that appeared to influence whether clinics adopted single or multiple interventions was their experience in diabetes QI; clinics with less experience were more likely to implement single interventions. Clinics in groups that focused on diabetes metrics but not in response to public reporting were more likely to implement multiple interventions. The authors asked quality directors from four doctor groups why clinics chose to operate multiple or single interventions in a given year. Their responses

indicated that clinics implementing single interventions were in the early stages of QI. The authors of the paper report the views of one quality director:

One quality director commented that, with the group's participation in the collaborative, its doctors were seeing standard comparative reports for the first time. The director said that these reports motivated clinicians to 'do something', but 'they just didn't have the bandwidth to do more'.

The authors also noted that clinics implementing multiple interventions sometimes did so in response to public reporting, but they were also often involved in externally sponsored QI projects. The authors report:

In one case, clinics had implemented a single physician-directed intervention as a 'first step' but the quality director of that group noted that 'we needed broader organisational change to sustain improvement'.²³⁰

Despite its methodological flaws, this study's findings suggest the public reporting of performance data alone does not stimulate sustainable organisational change. The public reporting of performance data is more likely to lead to sustainable QI if it occurs alongside other, large, externally supported QI initiatives. Single QI interventions implemented in isolation may not lead to sustainable QI. The study also suggests that QI occurs incrementally, and that the organisations may achieve sustainable improvement through a process of trial and error; single interventions may be a first step along this pathway.

Theory 8 summary

These studies suggest that clinicians who developed the WCHQ public reporting system were motivated by a desire to improve patient care, and retain control over how the system was developed and operationalised rather than to achieve market dominance. The indicators selected reflected clinical views of what constituted good care, and clinicians themselves were involved in the development of the case-mix adjustment algorithms. In terms of the theories under test, this involvement therefore both ensured that these data were valid and instilled a sense of ownership in these data which, in turn, made it much more difficult for practices to reject these data as invalid. When the programme was more widely implemented, Lamb *et al.*'s²²⁹ study indicated that clinicians did take steps to improve patient care, but responses were variable and no clinician group was able to respond to all of the indicators. Furthermore, Smith *et al.*'s²³⁰ study demonstrated that QI occurs incrementally and often requires support from other national QI initiatives to be successful. It also suggests that as practices gain more experience in QI, they are more able to implement more sustainable changes. This suggests that although clinical acceptance of the indicators as valid is more likely when the public report programme is led by clinicians, and this in turn means that clinicians are more likely to take steps to respond to such data, the success of QI initiatives also depends on the experience of the practice in implementing QI activities and the resources available to implement changes. This raises questions about what makes public reporting programmes more actionable and what support providers need to make sustainable changes. It is this theory that we turn to in the next section.

Theory 9: the degree to which performance data are 'actionable' influences providers' responses to the feedback of performance data

We now turn to testing theories which focus on what makes performance data actionable. In *Chapter 3*, we identified a number of theories which specified the conditions under which performance data may support or constrain attempts by recipients to take action in response to them. Again, it is important to note that programmes are not implemented in isolation but have to work alongside other initiatives that may support or inhibit their impact. Furthermore, programmes themselves are complex and embody a collection of different characteristics that may have a differential impact on whether or not their intended outcomes are achieved. In *Chapter 3*, we suggested that the following configuration of programme characteristics may influence the extent to which providers used data to initiate improvements to patient care.

- **Timeliness:** if data are not fed back to recipients in a timely way, they do not reflect current care and are less likely to be used as a catalyst for QI.

- Problem identification: performance data rarely provide a definitive ‘answer’ regarding the quality of care provided; rather, what leads to change is the discussion and investigation of the underlying cause of the level of performance indicated from the data.
- Nature of the indicator: process data are more useful than outcome data for QI purposes as they are better able to provide an indication about the cause of the poor outcomes or what needs to be improved.
- Level/specificity of feedback: performance data are more useful if they relate to individual clinicians or departments, as this enables action plans to be developed and implemented at ward level.

We explore the impact of these contextual configurations by reviewing the evidence on providers’ views of and responses to patient experience data because we hypothesise that these data *can* (but do not always) exemplify this configuration of programme characteristics. Patient experience data are a form of process data and, we can hypothesise, provide information on the providers’ performance on different dimensions of the care experience and therefore, give an indication of which care processes need to be improved. A number of initiatives that collect patient experience data, such as the GP patient survey in England, collect and feedback those data more frequently (biannually) and with a shorter time lag between data collection (6 months) and feed back than that for PROMs data collection, where there is often 1 year between data collection and feedback. It can also be reported at ward level as well as at the hospital level, so individual wards can compare how they perform with the hospital as a whole and, for national surveys, the national average. We start by reviewing studies that have examined whether or not the feedback of patient experience surveys has led to improvements in patients’ experiences. We also review studies that have explored providers’ views of patient experience surveys and their self-report of the QI initiatives that were undertaken. Finally, we consider studies that have examined the impact of interventions designed to support providers in responding to patient experience data.

Theory 9a: patient experience data are actionable and enable providers to take steps to improve patient care

Vingerhoets *et al.*²³¹

This study assessed whether or not the structured, individualised, benchmarked feedback of patient experiences to GPs in the Netherlands resulted in improvements to patients’ experiences of care. This study was conducted before national surveys of patient experiences were implemented in the Netherlands and therefore GPs in this study were unlikely to have had the same exposure to patient experience surveys as they do now. The sampling frame for the study was a sample of 700 GPs in the Netherlands stratified by level of urbanisation. From this sample, 60 GPs from 43 practices were recruited to the study and each was asked to recruit two cohorts of 100 consecutive attending patients, one at baseline and one 15 months later. Each cohort of patients was asked to complete a previously validated patient experience questionnaire covering nine dimensions of care. After matching for practice size, each practice was randomly allocated to a control arm and an intervention arm. The control arm practices received no feedback from the questionnaire. GPs in the intervention arm practices received a written 15-page report detailing the patient experience scores provided by their patient on each dimension, total scores and also reference figures for all GPs. The report also contained an abstract for a review of the determinant of patients’ evaluations of their care and a manual that explained how GPs might use the results of the survey. GPs in the intervention arm were also sent a questionnaire enquiring about any changes they had made to their own behaviour or the organisation of care. The results were analysed using multiple regression to examine if there were any statistically significant differences in patient experience scores between the two arms of the trial.

The authors found that, after controlling for the effect of baseline patient experience scores, patients’ evaluations of continuity of care and medical care were statistically significantly less positive in the intervention arm. There were no statistically significant differences in the other seven dimensions of patient experience, despite GPs reporting that they had made changes to their own behaviour or the organisation of care. The authors hypothesise that the lack of effect of their intervention may have been because the follow-up period was too short for any improvements to register, that GPs may have been too busy to implement changes and that the general shortage of GPs in the Netherlands may mean that GPs felt less

pressure to respond to feedback. They argue that the intervention may have been more effective if 'it is embedded in an educational programme or QI activity related to a specific clinical topic or group'.²³¹ They also suggest that the feedback may have functioned as a means of identifying specific topics for QI that needed to be explored in more detail before implementing specific QI activities.

This study was conducted in a different context from that experienced by current GPs, who had considerably more exposure to feedback from patient experience questionnaires. However, it does suggest that a 'one-off' feedback of patient experience data to GPs, without any public reporting or financial incentives attached to it, does not lead to improvements in patients' experiences of care.

Elliot *et al.*²²²

This study examined the feedback and public reporting of the Hospital CAHPS survey, which measures patient experiences and is publicly reported on a quarterly basis. The scheme was initially introduced by the CMS on a voluntary basis, and in 2008, 55% of eligible hospitals in the USA were involved in the scheme. However, CMS implemented a penalty of a reduction of 2% in the annual payment to hospitals if they failed to collect data from 2007 and failed to report it from 2009. By 2009, the percentage of hospitals participating in the scheme rose to 80%. Thus, the Hospital CAHPS scheme had financial incentives linked to hospital participation. It is perhaps also worth noting here that two of the authors who conducted the study worked for CMS, which was responsible for introducing the scheme.

The data from the survey are reported as the proportion of responses in the most positive categories (i.e. 'definitely yes', 'yes' or 'always') across nine domains measuring nurse communication, doctor communication, responsiveness of hospital staff, pain management, communication about medicines, cleanliness of hospital environment, quietness of hospital environment and discharge information. The authors adjusted these data for survey mode and patient characteristics. They compared scores on the survey between data published in March 2008 and March 2009 for 2774 hospitals in the USA that publicly reported data for both time periods to examine whether or not there had been any improvements in patient experience over time.

The authors found statistically significant but very small changes in patient experience scores between March 2008 and March 2009 on all nine domains except doctor communication. Most changes were < 1 percentage point difference in the scores in the top category between the two time periods. In their discussion the authors describe these changes as 'modest but meaningful improvements' and argue that their findings provide evidence that 'healthcare entities are able to use CAHPS feedback to improve patient experience'.²²² However, it is difficult to know the real significance of the changes for patients from these data and the study did not explore whether or how providers did use the data to improve care or what changes, if any, were made. Without a control group, it is difficult to know if such modest improvements in patient experience would also have occurred in the absence of any feedback. Furthermore, the time period over which the study was conducted was short and may not have captured the impact of any changes. In terms of the theory under test, this study suggests that, in the short term, the feedback and public reporting of patient experience data to providers leads to only very small gains in some domains of patient experience but not in doctor–patient communication. We now look at studies that have explored providers' views and attitudes towards patients experience data, and their reports of whether or not and how they used this feedback to initiate QI activities.

Barr *et al.*²³²

This study explored the impact of mandatory public reporting of patient experience on providers' QI activities. The study focused on the state-wide mandatory reporting of patient experience in Rhode Island, which was initially fed back privately to providers in 2000, and from 2001 onwards was fed back publicly. The 56-item survey was carried out annually on a random sample of patients discharged from each state licensed hospital in Rhode Island. The survey covered nine domains of patient experience: nursing care, doctor care, treatment results, patient education (including discharge information), comfort/cleanliness, admitting, other staff courtesy, food service and overall satisfaction/loyalty. The survey findings were

publicly reported as the hospital's score on each domain expressed as whether it was the same as, above or below the national average. Hospitals also received survey item data (expressed as percentage scores).

The sampling frame comprised four key executives in each of the 11 hospitals (CEO, medical director, nurse executive and patient satisfaction co-ordinator). Of the 52 positions identified, 42 people agreed to take part in the study (13 CEOs, eight medical directors, eight nurse directors and 13 patient satisfaction co-ordinators). The authors interviewed participants 1 year after the initial release of the first public report, either face to face or by telephone, and explored what QI activities had taken place in response to the patient experience survey.

The authors found that every hospital reported at least two QI initiatives within the domains reported in the survey. The most commonly reported areas in which improvement initiatives had taken place were admitting (nine hospitals), patient education (nine hospitals), nursing care (eight hospitals), treatment results (eight hospitals) and food service (eight hospitals). Less common areas were other staff courtesy (six hospitals), doctor care (five hospitals) and comfort cleanliness (four hospitals). Hospitals also reported being involved in other, broader QI initiatives, which could also have impacted on the domains reported in the patient satisfaction survey. However, the authors did not explore how the hospital's own score on these domains related to the QI efforts. The authors also found that although most hospitals had a decentralised approach for initiating QI initiatives, the reporting of the patient survey results was centralised and focused on senior management. Participants explained that they used the patient experience survey results to prioritise areas for improvement. They also noted that they had the greatest support for QI activities from the board and senior management and the least support from medical staff. Participants cited 'widespread support for QI, a culture and leadership fostering QI and a team approach'²³² as being important for successful QI activities.

This paper was a small-scale interview study in one location of the USA. The authors relied on hospital leaders' self-report of whether or not QI activities had taken place, which might have been subject to recall or social desirability bias. However, this paper provides another layer of evidence to suggest that some areas of patient experience were more likely than others to be subjected to QI efforts. It also suggests that senior managers played an important role in supporting QI activities.

Boyer *et al.*²³³

The study reports on providers' views and responses to a locally developed and implemented patient experience questionnaire for inpatients in a 2220-bed teaching hospital in France. The patient experience survey had been carried out yearly since 1998 and was ongoing at the time the paper was written (2006). It produced patient experience scores on a number of dimensions (medical information, relation with nurses, relations with doctors, living arrangements and health-care management) for the hospital as a whole and for each clinical department. The authors surveyed staff members in the hospital using a 26-item questionnaire which examined if staff had been informed about the overall hospital results, the results for their ward, how they were informed about the results, if the results were discussed, whether or not any action plans were developed as a result of the survey and staff attitudes towards patient experience surveys. The authors sent 502 questionnaires out to staff in the hospital, although they did not report what their sampling frame was or how it was determined. Of these, 261 (52%) of the questionnaires were returned.

The authors found that the specific results for the ward were less well known than the overall hospital results, with 60% of respondents indicating that they were aware of the specific ward results and 70% indicating that they were aware of the overall hospital results. However, 87% of staff indicated that they were more interested in the ward-specific results, compared with 13% who indicated that they were more interested in the overall hospital results. Respondents placed a higher value on open-ended comments than on standardised patient experience scores. Forty per cent of respondents indicated that the results of the patient satisfaction survey were discussed in staff meetings, 40% indicated that actions were taken to solve problems and 40% indicated that the survey had led to modifications to professional behaviour.

In their conclusions, the authors argue that one explanation for the insufficient use of the survey may be explained by 'a lack of quality management culture' and a lack of 'discussion of the results within the department'.²³³

This is a poor-quality study; the sampling frame for the study was not clear and the sample size and sample of participants who responded was small. The survey was conducted in one hospital and as such its findings may not be generalisable. Nonetheless, the study provides some indication that ward-level data were perceived as more useful than overall hospital performance. It also suggests that patient experience surveys are not a panacea to QI but that their use depends on the extent to which the data are disseminated within the hospital and whether or not they are discussed in ward meetings. It also implicates the importance of a broader, supportive hospital culture in facilitating the use of patient experience surveys to improve patient care.

Geissler *et al.*²³⁴

This study explored the motivators and barriers to doctors' use of patient experience data. The authors were particularly interested in doctors' views of the patient experience surveys distributed confidentially as part of the activities of the MHQP collaborative. This survey was conducted and fed back to clinicians every 2 years. However, they also explored doctors' views of other forms of patient experience data obtained from other sources. They developed a conceptual model to guide their investigation. They theorised that the degree to which doctor practices were engaged in initiatives to improve patient experience influenced the extent to which they make improvements in patient experiences. The degree to which doctor practices engaged in initiatives to improve patient experiences was influenced by organisational characteristics, such as culture, incentives, IT management and leadership, and by the characteristics of patient experience reports themselves, in terms of how they were disseminated, ease of use, timeliness and level at which the report was fed back. Here we focus on the findings relating to the nature of the reports themselves.

To test their model, Geissler *et al.*²³⁴ conducted 30-minute semistructured interviews with a sample of doctor groups in Massachusetts. The sampling frame was the 2007 MHQP state-wide doctor directory, with at least three doctors providing care to members of at least one of the five largest commercial health plans in Massachusetts, resulting in 117 doctor groups who were invited to participate. They interviewed leaders from 72 groups, giving a 62% response rate.

Their study did not specifically compare doctors' views of different types of patient experience surveys, but their findings do provide some insight into how the characteristics of the reports and the way they were fed back served to support or constrain the use of patient experience surveys in improving patient care. Participants indicated that they valued free-text responses from patients and sent positive responses to staff to boost morale, especially if individual staff were named. The negative ones were used to target particular departments or wards that were named in the feedback. This suggests that the free-text responses were valued because they allowed a more specific understanding of what was going well and what was not.

They also valued patient experience surveys, which provided support in interpreting and acting on the findings, such as those which provided a 'priority list' consisting of the 'ten most important areas or things that you could address that would have the biggest impact on improving patient satisfaction'.²³⁴ The timeliness of the data was also mentioned as important, with data provided on a frequent basis being seen as supporting efforts to improve care, and those with a large time-lag between data collecting and analysis being seen as less useful, as the following quotations from respondents illustrate:

This data has been more useful . . . because it's more timely. The data is available to us on an ongoing basis; we get it literally every day . . . so . . . the feedback is . . . more current.

I will get the MHQP and it's on stuff that happened a year and a half ago. That's very hard to go out to . . . practices and say we have got a . . . problem . . . you've got to do something about it . . . they say 'well that was a year and a half ago'.

Doctors also valued reports that provided data at the level of the individual clinician and were benchmarked against other groups' performance, so that they could compare their performance with that of others.

In terms of the theories under test, this study suggests that support with data interpretation, the level at which the data were reported and the timeliness with which the data were reported were seen by doctors as important in either constraining or supporting their efforts to improve patient experience.

Reeves and Seccombe²³⁵

This study aimed to explore providers' attitudes towards patient experience surveys and understand if and how they were being used in NHS hospitals. At the time of this study, annual patient experience surveys were conducted in specific patient populations: inpatients, emergency departments, outpatients and young patients (aged 0–18 years). Twenty-seven hospitals were purposively sampled from 169 NHS trusts providing acute care; the sampling frame was organised according to the size of the trust and whether they were inside or outside London. The person listed in the Health Care Commission records as being the lead for patient surveys was contacted to check whether they were the lead and, if so, invited to take part in an interview. It is not clear if interviews were undertaken face-to-face or by telephone. The interviews focused on views and uses of patient surveys, but they were not tape recorded and only notes were taken. This study therefore focuses on the views of those who lead in the trust on patient experience surveys and as such may not represent the views of frontline staff. As notes were taken, it is possible that key issues were missed and that participants' responses were filtered through the interviewer's frame of reference, leaving open possibilities of misinterpretation, misunderstanding of what the interviewee meant and selective listening or remembering. The study was funded by the HSCIC.

Participants drew attention to the trade-off between the timeliness and robustness of different sources of patient experience data. Participants noted that comment cards and suggestion boxes offered immediate feedback, and comments written on questionnaires were seen as useful in gaining the attention of clinicians and often provided details of incidents of poor care. As one participant commented 'Reading through the comments, even though our percentage scores are OK, you think "That shouldn't have happened"'.²³⁵ However, patient experience surveys were seen as more robust: 'Without a doubt, the national patient surveys are given the most weight. We have nothing else that is so sophisticated and would give us such useful data'.²³⁵ There appeared to be a distinction drawn by participants between 'soft' information, such as comments or complaints, and 'hard' evidence such as clinical or routine data. Patient experience surveys appeared to be seen as more 'robust' than comments and complaints.

Participants also commented about the methods through which survey findings were disseminated, most commonly through the organisation's intranet, newsletters, meetings where the contractors came into present findings and special events. In most organisations, the results were sent to senior staff who were expected to cascade the information down to junior staff, but participants reported that some groups of staff, such as doctors or more junior staff, were less likely to receive the results. They also commented on how difficult or easy it was to interpret those data; almost all participants felt that the Healthcare Commissions presentation of the published results were easy to interpret, especially the traffic light system, which shows whether the trust falls within the best 20% of trusts, the middle 60% or the bottom 20%. This helped trusts to 'see quite clearly where you are and where you should be'.²³⁵

However, when it came to acting on the findings from patient experience surveys, opinions were more varied. Some participants felt that feedback from patient experience surveys was not specific enough to be relevant to recipients, who, it was hoped, would act on the information. This was particularly seen as an issue for doctors, who participants felt were focused on their 'sphere of influence'. As one participant noted, 'The main criticism we have from doctors is "Make it specific to the area I work in and I will take

notice of it"'.²³⁵ They also noted variation in clinicians' 'receptiveness' to survey findings, with nurses perceived as being 'easier to engage' than doctors.

Almost all participants reported using patient survey results as the basis of action plans, and the authors give two examples of changes providers made. Both were in response to very specific issues highlighted by the survey: one was in response to the surveys highlighting problems with 'noise at night' that led to a range of efforts to reduce noise on the ward, and the other was variable information provided at discharge which led to changes in the way information was provided. Here, the surveys appeared to highlight problems with specific areas of care that were addressed. However, some participants also noted difficulties in formulating and then implementing action plans in response to the data. One participant commented that 'Just giving people the results doesn't mean they will take action. They need direction to make them do things and the frameworks to help them'.²³⁵ Other participants commented that they found it difficult to identify the reasons behind their successes or failures, and had difficulty knowing how to address shortcomings. Policy documents, published at the time of the study, promoted the idea of spreading best practice across the NHS, in line with theories of collaborative benchmarking. However, participants appeared divided in their enthusiasm for this idea. Some were interested in learning how others had made improvements, while others, in the authors' words, were 'not particularly enthusiastic' about identifying and learning from the best practices of others, although the exact form of their opposition is not reported.

In terms of the theories under test, this study indicates that providers were aware of the trade-off between timely and robust feedback and felt that both types needed to be integrated to provide a fuller picture of their performance. It also suggests that providers preferred feedback that was specific to their ward or 'sphere of influence' and that this was an important determinant of whether or not providers took action in response to this. However, it also suggests that providers needed support to identify the reasons behind their successes or failures and, in turn, to take steps to make improvements. This suggests that patient experience data do not always provide a clearer picture of the causes of good or poor care.

Boiko et al.²³⁶

This study explored primary care staff views and responses to the confidential feedback of a patient experience survey, similar to that used in the England GP patient survey. A random sample of 25 practices from Cornwall, Devon, Bristol, Bedfordshire, Cambridgeshire and North London agreed to take part in the study, and a random sample of their patients was mailed the patient experience survey. The practices received aggregate-level feedback for their practice, and each individual family doctor also received confidential feedback of their own scores on the patient experience survey. A purposively selected sample of 14 GP practices were then invited to take part in focus groups, which included 128 participants in all (40 GPs, 18 practice managers, 18 nurses, 20 receptionists, 13 administrators and 19 other staff members). The focus groups explored how practices had responded to the findings of the surveys, and also commented on two hypothetical situations in which some doctors in the practice received less favourable patient experience scores than other doctors.

Participants questioned whether or not patient experience surveys could adequately capture the 'complex reality of healthcare interactions' and contended that they focused on what was measurable to the exclusion of other important aspects of care. As one GP explained:

*A lot of this data that's collected in a measurable kind of way doesn't really represent reality. There's a kind of fixation on measurable outcomes but they don't really tell us what's going on.*²³⁶

Staff also drew attention to the trade-off between the increased relevance of local surveys that were less robust, versus the robustness of national surveys that were less specific to individual practitioners and did not include free-text comments. As one GP commented, 'We want to see data tailored to individual practitioner because we all practice [sic] differently'.²³⁶ Patient complaints were seen as more useful

because they allowed practices to understand where problems may lie. As one administrator noted, 'I think we learn a lot more from patients that write to us individually about complaints'.²³⁶

Participants also reported a number of changes they had made to services in response to the survey, including modifying their facilities, appointment systems and providing staff training. The changes largely related to organisational aspects of service delivery and operational matters. However, the authors commented that, for most practices, changes were 'rarely attributable directly to the survey feedback',²³⁶ rather, the survey had provided a 'nudge' to implement changes they were already considering. Participants mentioned a number of difficulties in responding to issues highlighted by the patient survey, including not having the resources to acquiesce to perceived unrealistic patient expectations (e.g. patients wanting the surgery to be open at weekends), balancing the sometimes conflicting demands of different groups of patients (e.g. some patients wanted music in the waiting room, whereas others did not) and the working patterns of GPs making it difficult to always fulfil patients' preferences. As one doctor summarised:

*Would you like the surgery to be open on Saturday? Yeah. Would you like us to go 24 hours? Yeah. Are you going to pay more taxes to have it open on Saturday? No.*²³⁶

They also felt that, even though they had made changes to the organisational aspects of the delivery of care, these had not been always been reflected in improved scores on patient experience measures which, as one GP described, were 'remarkably stubborn in terms of the change in perception by patients'.²³⁶

In particular, they acknowledged that it was very difficult to tackle an individual doctor's poor performance, especially when findings were fed back confidentially. It was only perhaps when these findings were shared more widely within the practice that change might occur. As one practice manager commented:

*If the survey results are between (the survey providers) and the doctor . . . there's absolutely no reason for them to change their ways is there? What is the motivation to change . . . ? It is only when this information becomes available to . . . the practice that things could start to change.*²³⁶

However, this respondent was also unsure exactly who in the practice could be expected to put pressure on 'poor performing' GPs to change their behaviour. Teams acknowledged the difficulties of having an 'unmanageable' GP in the practice but most teams indicated that they would support a doctor who consistently received poor patient feedback through mentoring, peer support sessions and interventions by a partner or manager. They also recognised that some GPs may not be 'a great communicator but they are great at doing something else'.²³⁶ Finally, staff felt that there was little external support for making changes in response to patient experience surveys. One GP complained that surveys had come out but that there was:

*very little support from anyone to say, right, this is how you can improve things that might help or we understand why you might be having problems . . . It has always been: here is your survey results, it is up to you now to sort it out.*²³⁶

The authors concluded that primary care staff view the role of patient experience surveys as serving a 'quality assurance' function, as they offered evidence that they were providing an acceptable standard of care. However, it was less clear that patient experience surveys fulfilled a QI function. Although patient experience surveys identified potential dimensions for change, 'actual changes were usually confined to "easy targets" for modification such as décor or playing music'.²³⁶ They note that 'issues such as the management of GPs with evidence of poor communication skills, or responding to other "interpersonal" aspects of care, were much harder to tackle'.²³⁶ They also argue that patient experience survey findings were only one of the 'spurs to action' to address problems that practices were often already aware of. In terms of the theories under test, this study suggests that although patient experience surveys may provide a clearer indication of areas of care that required improvement, there was no guarantee that this led to QI activities. The changes that were made focused on issues that staff were already aware of or on

the organisational aspects of service delivery, rather than on the 'harder to tackle' issues of communication skills and interpersonal behaviour.

Theory 9a summary

These studies suggest that the timeliness of some forms of patient experience data were valued by providers and that ward-specific data were perceived as more useful than higher-level hospital data. However, they also draw attention to the trade-off in characteristics of different forms of patient experience data. Qualitative or 'softer' data from comment cards, patient responses to open-ended questions or complaints were seen as providing data that were useful in providing a more detailed understanding of the nature and causes of problems but were seen as 'less robust' by providers, while patient experience surveys were perceived as focusing on measurable but less relevant aspects of patient care but were acknowledged to be more robust. Furthermore, providers questioned whether patient experience surveys were able to capture the real-life complexities of patient care. Providers felt that both sources were needed to provide a more rounded picture of patient experience.

However, although patient experience surveys identified potential dimensions for change, the studies reviewed here suggest that this did not always lead to steps to improve patient care. Furthermore, where changes did occur, these were not always directly in response to the findings of patient experience surveys, but reflect issues staff were already aware of from other data sources. The studies indicate a number of reasons for this. Some forms of patient experience data were perceived as not being specific enough to be 'in the sphere of influence' of certain clinicians. Providers were sometimes not clear on how to make changes and wanted guidance with this process. Others felt that they did not have the resources to meet patient demands, such as opening primary care services at the weekend. The studies suggested that, when changes were made, these tended to focus on addressing aspects of the organisation of care, or the so-called 'easy stuff', and that the more interpersonal aspects of patient care, relating to the behaviour and communication skills of individual clinicians, were much more difficult to address.

Theory 9b: making patient experience data more immediate and integrated into clinical discussions improves provider responses; theory 9c: providing tailored support to interpret and act on patient experience data improves providers' responses

We now examine studies that have attempted to provide additional support to providers in order to enhance the impact of the feedback of patient experience data. These have included making patient experience data more immediate and better integrated into clinical discussions and providing tailored support to help providers interpret data, identify problems with care and develop solutions.

Reeves *et al.*²³⁷

This study was designed as a pilot study to test the feasibility and impact of an intervention designed to improve the ways in which patient experience data were fed back to recipients. The intervention had a number of components, including (1) increasing the immediacy of the feedback, (2) providing specific feedback at ward level, (3) including patients' written comments in addition to the ward's scores and (4) an enhanced version of the intervention that also included ward meetings with nurses to discuss the findings of the survey and support them to act on the findings. It was hypothesised that the intervention would increase the likelihood that clinicians (in this case nurses) would take steps to improve patient care and, in turn, the patient experience scores for the ward would improve.

The study design was a RCT in two single site acute hospitals in London. All non-maternity inpatient wards were eligible for inclusion: 18 wards in trust A and 14 wards in trust B. Nine wards in each trust were randomly allocated to one of three arms, although the rationale for this number is not provided. Patient survey data were gathered in each trust specifically for the study, using the CQC's Inpatient Questionnaire collected through a postal questionnaire. A random sample of 160 patients discharged from each of the included wards over a 2-month period was taken at 4-month intervals during the study period (on six occasions in trust A and three in trust B) and they were mailed a questionnaire. The overall response rate during the study period was 47%.

In the control arm, survey findings were provided to the director of nursing in each trust, with no special efforts made to disseminate them to ward nurses. In the 'basic feedback' arm, individual letters were sent to nurses on the wards and their matrons which detailed (1) bar charts of scores on questions about nursing care, comparing the target wards scores with the scores for other wards in the arm and the national average; (2) as the study progressed, graphs of how the ward's scores had changed over time; and (3) transcription of patients' responses to a series of open questions on the patient experience survey. The 'feedback plus' arm received the letters and feedback, but ward managers were also asked to invite ward nurses to ward meetings with the researchers during working hours to discuss the survey findings. The main outcome measure of the study was the mean score of a subset of 20 questions from the CQC inpatient questionnaire, which was termed the Nursing Care Score. The authors do not provide details on why and how these questions were chosen or on whether or not the scale was psychometrically valid and tapped into a common factor of nursing care. Notes were taken at meetings with nurses in the feedback plus arm and ward managers in the basic feedback arms were also contacted to establish whether or not actions had been taken in response to the survey, although details of how this occurred were not provided.

Multiple-level regression was used to examine changes in the Nursing Care Score in wards in each arm. This analysis suggested that there were no statistically significant differences in the changes in ward patient experience scores over time between the control arm and the basic feedback arm. The changes in the Nursing Care Score from baseline to follow-up for the wards in both basic and feedback arms were negative in all wards except two, suggesting that their patient experience scores had worsened over time. The authors report that, when asked, 'none of the basic feedback ward managers identified specific actions resulted from the printed results'. This suggests that improving the timeliness of feedback and providing ward-level specific feedback alone was not sufficient to lead to improvements in patient experience scores. There was a statistically significant difference in the changes in scores over time in the feedback plus arm compared with the other two arms, with scores improving over time. However, there was much more variation in changes in the Nursing Care Score from baseline to follow-up in the intervention arm, with some wards staying virtually the same (three wards with changes < 1.0 point either way), some worsening (two wards) and one improving. Graphs showing the aggregated rate of improvement in the three groups also demonstrate that the differences between the three groups was largely due to the fact that control group and basic feedback patient experience scores worsened over the study period, while a very small overall improvement was observed in the intervention arm.

The authors reported some of their own impressions from ward meetings to understand the function of the ward meetings and explain why they lead to improvements in patient experience scores. Attendance at meetings was variable, but the attendance of matrons 'had a positive influence', as they offered 'suggestions for improvement', encouraged 'ward staff and nurses to take responsibility for results' or supported 'efforts to implement changes'.²³⁷ The authors felt that the ward meetings facilitated 'nurses engagement' and noted that the written comments in particular stimulated the nurses' interest. The authors also noted that staff needed prompting from them to focus on understanding the patient feedback and planning strategies for improvement; without this staff were more 'inclined to discuss . . . the many difficulties they experienced in fulfilling their duties; staffing shortages, NHS policy or their perceptions of hospital managers'.²³⁷ We can hypothesise that the discussion of more general matters either was a lapse of focus on the nurses' part or served to explain or justify the difficulties in understanding and acting on the data.

The verbatim quotations that are offered by the authors suggest that nurses felt that some of the survey findings had a rational explanation and did not constitute an indicator of poor care; for example, they explained that 'Patients think we are talking in front of them as they are not there because we have to talk quietly to maintain confidentiality'.²³⁷ They also appeared to implicate patients themselves in making unreasonable demands by using 'call buttons for trivial reasons' so that 'it would not be good use of our time to answer them all immediately', suggesting, again, that they questioned if the survey indicators tapped into 'good care'.²³⁷ The authors also note that it was 'difficult to ascertain clear examples of innovations in patient care' as a result of the patient survey feedback and the meetings. The most common responses were that

nurse managers raised the issues in ward meetings and handovers and 'reminded nurses of the importance of fulfilling their duties relating to ensuring patients' experiences were positive'.²³⁷

This study was a pilot study and, as such, has a number of methodological shortcomings. The randomisation process was not masked, leaving open the possibility that the research team's preferences or knowledge of the wards influenced the allocation. It was not a cluster RCT; wards in the same hospital were randomised to different arms, leaving open the possibility that contamination between the three arms occurred, diluting the impact of the intervention. It is not clear whether or not the changes in patient experience scores in the intervention arm were statistically significant over time, only that they were statistically significantly different from the changes that occurred in the other two arms. The qualitative data were not collected systematically and many of the insights into the meetings were derived from the researcher's impressions, with few verbatim quotations from participants.

Despite these limitations, the study does provide some lessons about the elements of the feedback process that may support or constrain the actions taken in response to the feedback of patient experience data. In terms of the theory under test, this study suggests that improving the timeliness of feedback and providing ward-level-specific feedback alone was not sufficient to lead to improvements in patient experience. Indeed, under these conditions, patient experience scores worsened. The facilitated meetings did lead to small improvements in patient experience as measured by patient experience scores, but it is not clear if these in themselves were statistically significant. These meetings functioned as an opportunity for nurses to air their concerns about the data (and their general working conditions) and as a means of raising nurses' awareness about the data, rather than as sites for the strategic planning of improvements to patient care.

Davies *et al.*²³⁸

This study evaluated the impact of a peer and researcher support intervention coupled with a modified process of collecting and feeding back patient experience survey data on the QI activities of a small number of providers. Nine providers who had previously expressed an interest in learning to use patient experience data more effectively were invited to join a collaborative. One group left the collaborative early on in the project, leaving eight groups involved. They were invited to submit suggestions for how the current CAHPS survey could be modified to make it more useful for QI purposes, and the research team used the suggestions to refine the survey. The resulting survey covered five domains: scheduling and visit flow, access, communication and interpersonal care, preventative care and integration of care. It also included a question that measured a global rating of care, a question on whether or not patients would recommend the service to others and an open-ended question. The survey was administered to a random sample of recently discharged patients and fed back to providers on a rolling monthly basis to provide 'real time' data to providers, in the hope that this would 'support a rapid cycle of quality improvement'.²³⁸

In addition, the providers participated in a patient experience 'action group', which followed a 'model of collaborative learning'.²³⁸ This involved a group leader and team members attending a full-day bimonthly meeting facilitated by QI advisors, which focused on supporting the groups to understand and interpret the survey data, prioritise areas for improvement, set targets and action plans to address any issues raised by the patient experience data, and monitor their progress in implementing these plans. The impact of the intervention was evaluated using a mixed-methods cohort study, including measuring changes in patient experience scores during the 18-month intervention and interviews with collaborative methods to explore their experiences. All eight leaders participated in initial interview but, by the time of the follow-up survey, two groups had left the collaborative and two leaders had changed positions. Consequently, the six original leaders and one new leader were interviewed at follow-up. In addition, four leaders invited other members of their team to take part in interviews.

At follow-up, six groups had used at least one of the suggested tools to explore the possible reasons for their patient experience survey results. These tools included walkthroughs (five groups), patient interviews (two groups), patient focus groups (two groups) and cycle time surveys (one group). Four groups had used one of the suggested interventions designed to improve patient care, including scripting for clinic staff

(two groups), communication skills training (one group) and patient education materials. Two groups had developed their own interventions. The leaders reported feeling that the group's support had been useful in creating momentum and motivation to implement the changes, and had provided an opportunity to learn from others. As one member commented:

*It lets you know that you are not alone . . . we tend to blame our workers if we get bad outcomes but if the whole world is getting bad outcomes . . . perhaps . . . it's a common culture. We all have stories about success and failure and sharing those stories is helpful.*²³⁸

Six leaders decided on and attempted to implement at least one intervention to improve patient care; four leaders reported that they had implemented the intervention as originally planned, while two reported problems with implementation. These latter two groups reported that other organisational changes had competed for priority and they had focused on those instead. For the other four groups, however, the impact of these interventions was not always reflected in follow-up patient experience survey results. Many of the changes in the scores were not statistically significant and were sometimes in domains that were not directly related to the focus of the intervention that had been implemented. Of the four groups who had implemented an intervention, 'three had some change in the direction they had hoped for'.²³⁸ For example, one team had attempted to improve communication between staff and to patients about waiting times and test results, and found that more patients reported feeling that they had been kept informed about this in the last 6 months.

Two groups showed 'mixed or negative effects'. One group had delivered communication skills training to staff and found a 'slight improvement' in patients reporting that doctors explained things to them in a way that was easy to understand, but a decline in the percentage of patients reporting that doctors spent enough time with them. Staff reported that some aspects of the communication skills package were perceived as conflicting with the group culture in suggesting that the problem lay with individual clinicians who need to work on their skills, rather than working together as team. Another had tried to reorganise their clinic to improve visit flow but found that patients reported longer waiting times and not being informed about their wait. This was perceived to be because patients arrived earlier than they were asked to and, therefore, perceived this as an additional delay. Thus, attempts to improve one element of care seemed to have a detrimental impact on other aspects. Following a post hoc analysis of their findings, the authors identified that 'the two groups that succeeded most clearly in improving their patient experience worked on interventions that required no major change in clinician behaviour specifically for the project' and which 'aimed for modest improvements that did not require complex changes'.²³⁸

This was a small case study of six doctor practices that were self-selected, highly motivated to take part in the study and relatively experienced in improvement activities. As such, the findings of this study may not be generalisable to other practices that were less motivated and less experienced. However, it provides some valuable lessons as to the circumstances under which the feedback of patient experience data does and does not lead to the successful implementation of QI strategies. It suggests that teams need considerable support to interpret and understand patient experience data, and that they need to conduct more specific investigations to identify the cause of negative patient experiences. Furthermore, it indicates that implementing change is challenging, and that the more complex the issue, the more challenging it is to effect change. Those that succeeded in this study attempted relatively simple interventions that did not require substantial changes to clinical behaviour. Finally, it also demonstrates that change to one aspect of patient care can have a detrimental impact on other aspects of the systems. Thus, as the authors argue in their conclusion, it is possible to produce small improvements in patient experience by making changes to simple aspects of patient care, but 'it is difficult to . . . leverage more substantial change without a more comprehensive strategy that is organisation-wide and regarded as fundamental to organisational success'.²³⁸

Theories 9b and 9c summary

Here we have reviewed two different interventions that aimed to increase the impact of the feedback of patient experience data on improvements to the quality of patient care. They suggest that enhancing

the feedback alone, through providing timely, ward-level data to nurses, is not sufficient to lead to the implementation of QI strategies in response to these data. Supplementing this feedback with ward-level meetings served an important function of addressing nurses' concerns about the validity of the data, raising awareness of the data and reminding nurses about their role in supporting positive patient experiences. However, these meetings did not serve as opportunities for the strategic planning of QI activities and led to only small improvements in patient experiences.

The other study evaluated an intervention that focused on hospital leaders who were involved in modifying the patient experience questionnaire used and received regular, timely feedback of patient experience data and both expert and peer support to interpret the data, investigate the causes of poor care and implement interventions to address any issues. Despite this considerable amount of support, the findings from this study suggest that some improvements in patient experience occurred but were hard-won. Those that succeeded focused on simple interventions that did not require complex changes or changes to clinical behaviour. This reinforces the findings from the studies reviewed previously in this section that changes to the behaviour of clinicians are more difficult to achieve. The study also demonstrated that changes to one aspect of patient care can have unintended effects on other aspects. The lesson from this study is that significant and sustained improvement in patient experience in response to feedback can only be achieved with system- and organisation-wide strategies.

Chapter summary

In this chapter we have explored the ways in which different contextual configurations influenced the mechanisms through which providers respond to the feedback and public reporting of performance data and the resulting outcomes. We have tested three main theories:

- theory 7: financial incentives and sanctions influence providers' responses to the feedback of performance data
- theory 8: the perceived credibility of performance data influences providers' responses to the feedback of performance data
- theory 9: the degree to which performance data are 'actionable' influences providers' responses to the feedback of performance data.

Theory 7

The findings of our synthesis suggest that greater improvements in the quality of patient care occur when providers are subjected to both financial incentives and public reporting than when they are subjected to either initiative alone.^{215–217} We also found that the feedback of performance indicators to providers who are subjected to neither public reporting nor financial incentives rarely led to formal or sustained attempts to improve the quality of patient care, particularly when providers themselves did not trust the indicators.^{219–221} Under these conditions, the feedback of performance was more likely to lead to providers improving the recording and coding of data, which may be an important first step in increasing their trust in the data themselves and also providing a basis from which further QI initiatives may occur.

However, we also found that financial incentives have only a short-term impact on QI if they are used to incentivise activities that providers already perform well in and when providers reach the threshold at which they would receive the maximum amount of remuneration.²¹⁸ Furthermore, we also found quantitative²¹⁸ and qualitative evidence¹⁸ to indicate that financial incentives, together with public reporting, may lead to 'tunnel vision' or effort substitution, that is, focusing on aspects of care that are incentivised to the detriment of care that is not, especially when providers do not feel that the indicators adequately capture quality of care. There is also evidence to suggest that when providers are subjected to both public reporting and financial incentives attached to these indicators, but they do not feel that the indicators are valid or contribute to patient care, this can lead to the manipulation or gaming of the data.^{18,80,81} This is not necessarily the result of active attempts to 'cheat' the system on the part of providers. Rather, the use of financial rewards can create

perverse incentives that are at odds with the inherent clinical uncertainty of conditions such as depression. Under these conditions, clinicians have to find a way to manage this clinical uncertainty, while at the same time ensuring that they are not financially penalised for doing so.

Theory 8

Our synthesis suggests that adequate case-mix adjustment and the accurate coding and recording of data were essential for providers to have any trust in performance data.^{97,195,224} Both the source of performance data and the process through which they are collected and presented are important influences on whether or not such data are perceived as credible by clinicians. We also found support for the theory that clinicians perceived data from patients' notes as being more credible than performance data derived from administrative data.^{225–227} However, our synthesis also indicated that clinical involvement in the design of the public reporting initiatives was a better explanation of their success, or otherwise, than the nature of the data alone. We therefore tested the theory that providers respond differently to public reporting initiatives that are imposed on a mandatory basis by national or state governments or regulatory authorities compared with those that are led by clinicians.

We found that mandatory public reporting systems were perceived by providers to be governed by political motives, rather than by a desire to improve the quality of patient care.²²³ Mandatory public reporting systems focused the attention of hospital leaders and frontline staff on quality issues, particularly for those who had no previous experience of QI activities.²²⁸ However, unless the issues raised were also identified as a problem based on clinicians' day-to-day experience, by their peers or by the internal data collected by the providers themselves, mandatory public reporting systems were perceived as leading to 'tunnel vision'.⁹⁸ In particular, for those who were already engaged in QI activities, mandatory public reporting systems could artificially focus their attention on a limited number of issues, deemed important by government or regulatory bodies, at the expense of other clinically important areas.²²⁸ Furthermore, although scoring poorly on an indicator included in a mandatory public reporting system could 'kick start' a response from providers, it was only through analysing internally collected data that providers could understand the source of the problem and identify possible solution to rectify this.⁹⁸

We also found that clinicians engaged with clinician-led public reporting systems because they supported their primary motive of improving the quality of care through comparing performance and sharing best practices.¹⁹⁹ The perceived mechanisms of change were those evoked by audit and feedback and benchmarking theories; that is, clinicians had an intrinsic motivation to do a good job and were prompted to improve if their performance was poorer than they would like or poorer than that of their peers. This engagement in turn led to the development of indicators that reflected the information that clinicians needed to improve patient care. As clinicians were involved in the design of the indicators, it was difficult for them to then dismiss the data as invalid.

However, we also found that although clinical acceptance of the indicators as valid was more likely when the public report programme was led by clinicians, and this in turn meant that clinicians were more likely to take steps to respond to such data, the success of QI initiatives also depended on the experience of the practice in implementing QI activities and the resources available to implement changes.²³⁰ Providers reported that considerable resources were required to enable them to respond to issues highlighted by the feedback of performance data. Those who are already engaged in QI activities may have been more likely to set up internal data processes that inform QI on an ongoing basis and, thus, were better able to respond when external data shone a light on poor performance.

Theory 9

We tested the theory that the feedback of patient experience data can (but do not always) embody a cluster of characteristics that render it easier for providers to use these data to initiate improvements in patient care. These include that idea that patient experience data provide a clearer indication of which care processes need to be improved, can be fed back in a timely manner and can be reported at ward as well as provider level. We found that the timeliness of some forms of patient experience data were valued by providers and

that ward-specific data were perceived as more useful than higher-level hospital data.²³³ Our synthesis also highlighted the trade-off in characteristics of different forms of patient experience data. Qualitative or 'softer' data from comment cards, patient response to open-ended questions or complaints were seen as providing data that were useful in providing a more detailed understanding of the nature and causes of problems, but were seen as 'less robust' by providers, while patient experience surveys were perceived as focusing on measurable but less relevant aspects of patient care, but were acknowledged to be more robust. Providers felt that both sources were needed to provide a more rounded picture of patient experience.²³⁵

However, although patient experience surveys identified potential dimensions for change, we found that this did not always lead to substantial improvements in patient care.²²² When changes did occur, these were not always directly in response to the findings of patient experience surveys but reflected issues staff were already aware of from other data sources, a similar finding to our synthesis of other forms of performance data.²³⁶ We identified a number of reasons for this, which were also similar to those identified in our synthesis of studies addressing the credibility of other forms of performance data. Providers were sometimes not clear on how to make changes and wanted guidance with this process.²³⁵ Others felt that they did not have the resources to meet patient demands, such as opening primary care services at the weekend.²³⁶ Our synthesis also suggested that, where changes were made, these tended to focus on addressing aspects of the organisation of care, or the so-called 'easy stuff', and that the more interpersonal aspects of patient care, relating to the behaviour and communication skills of individual clinicians, were much more difficult to address.^{222,232,236,237} Furthermore, when we reviewed studies evaluating interventions designed to support providers to interpret the data, investigate the causes of poor care and implement changes, we found that those that succeeded focused on simple interventions that did not require complex changes or changes to clinical behaviour.²³⁸ Our synthesis also indicated that changes to one aspect of patient care can have unintended effects on other aspects and that significant and sustained improvement in patient care in response to feedback can only be achieved with system, organisation-wide strategies. This conclusion is shared by other realist syntheses evaluating other complex organisational interventions.⁸⁹

Chapter 6 Review methodology: feedback of individual patient-reported outcome measures in the care of individual patients

Searching for and identifying programme theories

In this chapter, we describe the process through which we conducted our realist synthesis of individual-level PROMs feedback in the care of individual patients. As discussed in *Chapter 2*, we conducted one search for programme theories for PROMs feedback at both the aggregate and the individual level. Details of these searches are provided in *Chapter 2* (see *Searching for and identifying programme theories*). The search strategy can be found in *Appendix 1*.

JG and SD screened the titles and abstracts of the 748 retrieved references to identify potentially relevant papers according to the following criteria.

Inclusion criteria

- The paper provides a theoretical framework that describes how individual PROMs feedback is intended to work.
- The paper provides a critique, review or discussion of the ideas underlying how individual PROMs feedback is intended to work.
- The paper provides stakeholder accounts or opinions of how individual PROMs feedback does/does not work.
- The paper outlines, discusses or reviews potential unintended consequences of individual PROMs feedback.

Exclusion criteria

- The paper reports findings in which a PROM is used as a research tool (e.g. an evaluation of an intervention or a study exploring the HRQoL of specific populations).
- The paper is focused on evaluating the psychometric properties of a PROM.
- The paper reviews the psychometric properties of a PROM or collection of PROMs.
- The paper provides advice or recommendations for which PROM to use in a research context.

An initial screen of the titles of these papers identified 111 papers for potential inclusion. Screening the abstracts identified 47 papers for inclusion, of which 21 contributed to the final synthesis. Citation tracking of these papers and additional searches identified a further 18 papers that contributed to the identification of candidate programme theories. A total of 39 papers were included in programme theory elicitation process. *Figure 16* summarises the process.

Focusing the review and selecting programme theories

The process of cataloguing the different programme theories underlying PROMs feedback at the individual level (reported in *Chapter 7* of this report) allowed us to identify the inner workings of these interventions as perceived by those who design, implement and receive these interventions. The focus of our review was agreed by the project group in an iterative manner over a series of meetings, taking into account the issues raised by our patient group and an initial 1-day workshop with a group of stakeholders. At the

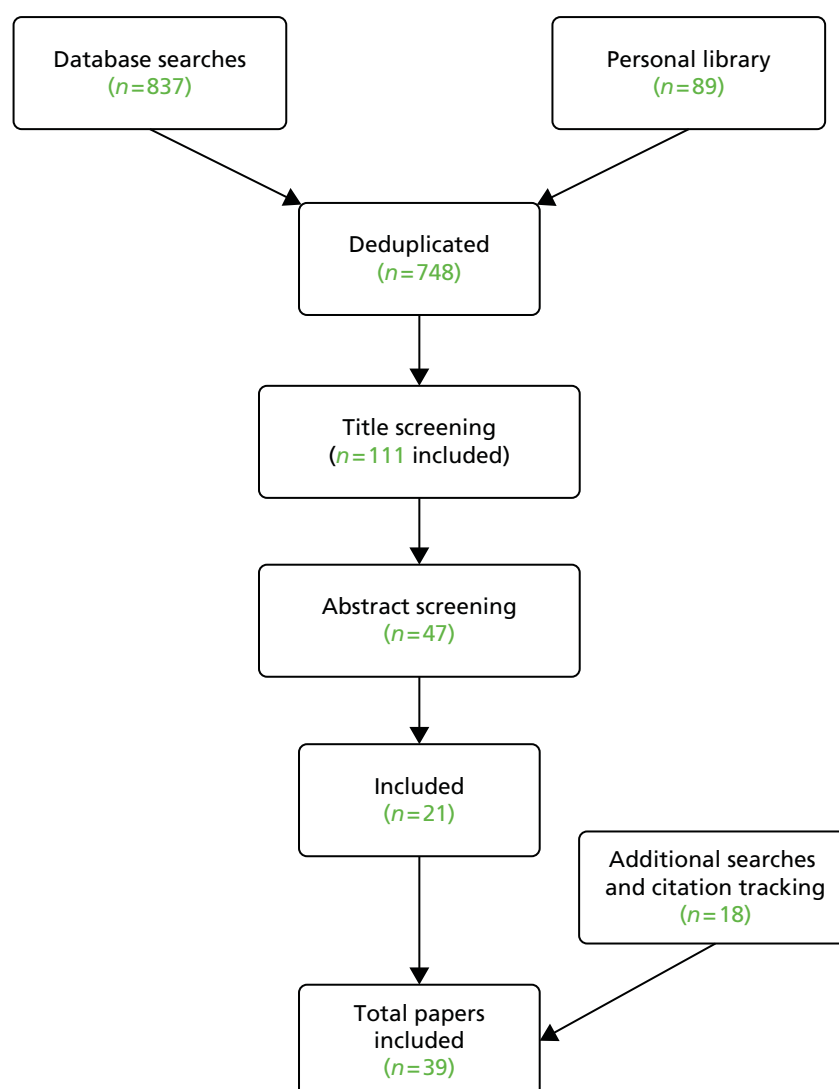


FIGURE 16 Theory searches for individual PROMs feedback.

initial stakeholder workshop and initial patient group meeting, we presented our initial programme theories and a basic logic model of the feedback of individual PROMs. Our patient group consisted of three 'expert' patients: one was a retired GP, one had previously worked for the NHS Commissioning Board and the third worked for a national charity, Arthritis UK. Our stakeholder event included the following stakeholders:

- three analysts on the national PROMs programme from NHS England
- an analyst working on the national PROMs programme from the HSCIC
- Matron for Surgery, Anaesthesia and Theatre
- Senior Sister for Surgical Pre-Assessment
- Director of Operations at a NHS trust
- Representative from the Royal College of Nursing with expertise in PROMs
- a consultant surgeon
- two academics with expertise in orthopaedics and PROMs
- two patient representatives.

We presented our initial programme theories at these meetings and invited participants to comment on these ideas and to refine, extend and prioritise them. During this meeting, we found that stakeholders focused more on discussing and extending our programme theories than they did on prioritising our

theories. Key ideas raised at this meeting regarding the use of PROMs in the care of individual patients included:

- At the individual level, PROMs could be a starting point to be used in chronic illness management. However, regular completion of PROMs requires matched IT software (namely computer data entry from the patient that can give immediate feedback to the patient and the clinician).
- The importance of PROMs for patients is that they are assumed to standardise assessment and history taking and allow a 'shared language' to be used between the clinician and the patient.
- In the future, PROMs could be used like a blood test or X-ray, as another tool to aid clinical decision-making. A PROM is used as it allows the clinician to gather all of the evidence systematically.

At our patient group meeting, we also found that the group spent more time discussing and commenting on our programme theories than they did prioritising them. The issues raised in our patient group included:

How patients might use patient-reported outcome measures

- It can make patients more willing to start a conversation with providers and other services, making them feel more comfortable in asking questions, thus empowering patients in their relationship with providers.
- Completing a PROM questionnaire can make the patient focus more on particular symptoms. This could be useful for self-awareness, although it may also mean that patients dwell on symptoms that did not previously worry them. For some individuals, it is best not to be so informed.
- With a system of patient–clinician partnership, PROMs should be empowering to the patient and useful to the clinician.

How clinicians might use patient-reported outcome measures

- Idea of patients using PROMs in dialogue with clinicians: 'clinical team' may be a better phrase, to show that this is understood broadly.
- Those clinicians who are already aware of looking at patient needs may be the ones who make most use of PROMs.
- It may be possible to support more effective dialogue with patients without the added paperwork of PROMs. However, having the PROMs form perhaps helps to ensure that this dialogue happens.

These meetings were undoubtedly useful in helping the project team identify the programme theories underlying the use of PROMs in the care individual patients. However, owing to the structure of the meetings as an open discussion, they offered less scope for stakeholders to prioritise our programme theories and determine the focus of the review. We would suggest that future reviews consider supplementing open group discussions with the use of the nominal group technique^{239,240} in order to involve stakeholders determining the focus of the review.

Following the stakeholder workshop and patient group meeting, we held a project team meeting to reflect on the issues raised. At this stage, we agreed to focus the individual-level synthesis on how the feedback of individual PROMs data may support the care of people with long-term conditions. We agreed to focus on their role in supporting personalised care planning. During subsequent project group meetings and a further patient group meeting, we discussed a number of possible ways that we could structure the review. These included the following.

A review to compare patient-reported outcome measures use in different long-term conditions

With this approach, we would compare the process and outcome studies between different long-term conditions that represent different contextual conditions, for example conditions where the focus is largely on monitoring more 'clinical' indicators (e.g. diabetes and cancer) but where disease management may

also have psychosocial impact, compared with conditions where monitoring psychosocial issues is the main focus (e.g. mental health and psychotherapy). Another set of contextual factors includes conditions for which routine outcome measurement has been mandated to support both quality assurance and the care of individual patients (e.g. mental health) and those for which it has not or where different data are collected for different purposes (e.g. the National Cancer Survey is collected purely for quality assurance purposes and not to support the care of individual patients).

A review of specific blockages in the patient-reported outcome measures feedback implementation chain

With this approach, we would focus on examining a specific part(s) of the implementation chain from intervention to expected outcomes in detail. This implementation chain runs from PROMs completion by patients, feedback to clinicians, improvements in communication, the detection and discussion of patient problems, actions taken to address them and improved patient care and outcomes. The goal would be to answer the question 'what are the key blockages to PROMs feedback improving the care of patients with long-term conditions and through what processes and in what circumstances can they be overcome?' Key areas to explore are whether or not and how PROMs feedback might support patients to raise concerns with clinicians, how PROMs data might prompt clinicians to raise issues with patients and if any subsequent action is undertaken in response.

A review of the 'system strains' to patient-reported outcome measures feedback

With this approach we would identify the 'system strains' to PROMs feedback improving patient care and explore how they operate, and whether or not and how the purported solutions resolve them. There are a number of different 'pinch points' along the implementation chain from PROMs feedback to improved patient outcomes. We identified five of these 'system strains'.

- System strain 1: standardised PROMs may not capture patients' experiences effectively, so it has been suggested that individualised PROMs are more appropriate for use in individual patient care. In what circumstances and through what processes might individualised and standardised PROMs support or constrain communication during the consultation?
- System strain 2: there is a tension between having long instruments that are psychometrically valid and comprehensive, and having those feasible in clinical practice where there is a need for shorter instruments that patients can easily complete and can be easily scored. A solution to this is the use of item banks with fewer items that are more relevant to patients [e.g. Patient Reported Outcome Measurement Information System (PROMIS®)]. What is the evidence that this system is more feasible for use in routine patient care than standard approaches?
- System strain 3: it is difficult to integrate pen-and-paper questionnaires into patients' notes and into the workflow of the clinic; therefore, PROMs are now collected electronically and integrated into a patient's electronic health record. What is the evidence that this means PROMs are more easily accessible in the clinic and are therefore more likely to influence clinical decision-making?
- System strain 4: there is a concern that clinicians do not know how to address the issues identified by PROMs feedback. One suggested solution is to incorporate management guidelines into the feedback that advise clinicians and/or the patient what to do if the patient's score on y subscale is above x. When, how, why and for whom does this solve the problem?
- System strain 5: our patient group also identified another system strain, which was also evident in the PROMs programme theories and linked to system strain 1 above (reported in *Chapter 7*); although PROMs feedback may enable patients to raise issues with clinicians or prompt clinicians to discuss issues with patients, there is also the risk that PROMs feedback may detract from the relationship-building process.

We considered all of the options above and agreed to focus our review on a combination of these approaches. A number of reviews had already focused on the practical or logistical aspects of PROMs implementation.^{59,60} Furthermore, a review has recently been published exploring electronic patient-reported outcomes (ePROs). We decided to focus our review on exploring system strains 1 (individualised vs.

standardised measures) and 5 (constraining vs. supporting the clinician patient relationship). The precise focus of the review evolved over two project group meetings and a pilot synthesis (described below). We aimed to explore these strains in relation to particular points on the implementation chain: (1) how PROMs support patients to raise or share issues with clinicians and (2) how PROMs support clinicians to discuss issues with patients and subsequently to take action to address these issues. To do this, we decided to compare studies explore PROMs in primary and secondary care mental health settings and PROMs use in oncology and palliative care settings, to provide contrasting contextual conditions. Therefore, the objectives of our review were to:

- understand the circumstances in which and processes through which PROMs feedback enables patients to share concerns with patients
- understand the circumstances in which and processes through which PROMs feedback raises clinicians' awareness of patient problems and prompts discussion of these issues during the consultation.

Searching for empirical evidence and selection of studies

From March to May 2015 we searched for empirical evidence for the impact of individual PROMs feedback, initially by forwards and backwards citation searches of six key papers^{47,48,58,60,62,65,241} in Web of Science Core Collection: Citation Indexes (Thomson Reuters) from 1900 to present.

The forward citation searches for Boyce *et al.*,⁵⁸ Cowley *et al.*,²⁴¹ and Knaup *et al.*⁴⁷ found 695 records, reduced to 442 when the duplicates were removed. The forward citation searches for Valderas *et al.*⁴⁸ found 113 citations, for Greenhalgh *et al.* 2005⁶² found 124 citations and for Greenhalgh *et al.* 2013⁶⁵ found seven citations, plus a further five that cited the previous seven. In total, this produced 605 references. In addition, the references lists of five purposively selected systematic reviews were also reviewed. These were (number of references in each review appear in brackets) Valderas *et al.*⁴⁸ ($n = 65$), Chen *et al.*⁴³ ($n = 66$), Greenhalgh *et al.*^{62,65} ($n = 59 + 41$), Boyce *et al.*⁵⁸ ($n = 71$) and Krageloh *et al.*⁴¹ ($n = 70$). Both the forward citation tracking references ($n = 605$) and the backward citation references lists ($n = 372$, without deduplication) were screened by JG and EG according to the following inclusion and exclusion criteria.

Inclusion criteria

The study contributes to explaining:

- how PROMs may support patients to raise issues with clinicians
- how PROMs may support the relationship-building process between patient and clinician
- how PROMs may constrain the relationship-building process
- how PROMs may support discussion during the consultation.

We also included studies exploring:

- clinicians' and patients' experiences of using PROMs feedback in oncology, palliative care, and primary and secondary mental health settings
- patients' experiences of completing standardised and individualised PROMs.

Exclusion criteria

- The study does not add anything or contribute to theory testing and refinement.
- The study explores the psychometric properties of PROMs.
- The paper reports findings in which a PROM is used as a research tool (e.g. an evaluation of an intervention, a study exploring the HRQL of specific populations).
- The paper is focused on evaluating the psychometric properties of a PROM.
- The paper reviews the psychometric properties of a PROM or collection of PROMs.

From the forward citation searches, 130 were included after reviewing the titles and 15 were included after reviewing the abstracts. After a full-text review of these 15 papers, eight studies contributed to the final synthesis. From the backward citation tracking/review of references lists, 29 were included following a review of the titles. After full-text review of these 29 papers, 19 studies contributed to the final synthesis. In addition, during the synthesis, the references lists of included papers were checked and key author searches (e.g. Velikova, Wolpert and Dowrick) were undertaken in MEDLINE (July–September 2015) and a further nine papers were identified and included in the synthesis. *Figure 17* summarises this process.

Data extraction, quality assessment and synthesis

This was an iterative process undertaken by JG, KG and EG with feedback from members of the wider project group (NB, RP and CV). Study selection, data extraction, quality assessment and synthesis, plus additional literature searching, occurred simultaneously. As discussed in *Chapter 1*, studies were selected on the basis of their contribution to theory testing. In some instances, the whole study contributed to theory testing and in others only a fragment or fragments of the study were relevant to the theory. Each fragment of evidence was appraised, as it was extracted, for its relevance to theory testing and the rigour with which it had been produced.⁸⁷ Therefore, quality appraisal related specifically to the validity of the causal claims made in these subset of findings, rather than the study as a whole. Trust in these causal claims is also enhanced by the accumulation of evidence from a number of different studies that provide further lateral support for the theory being tested.

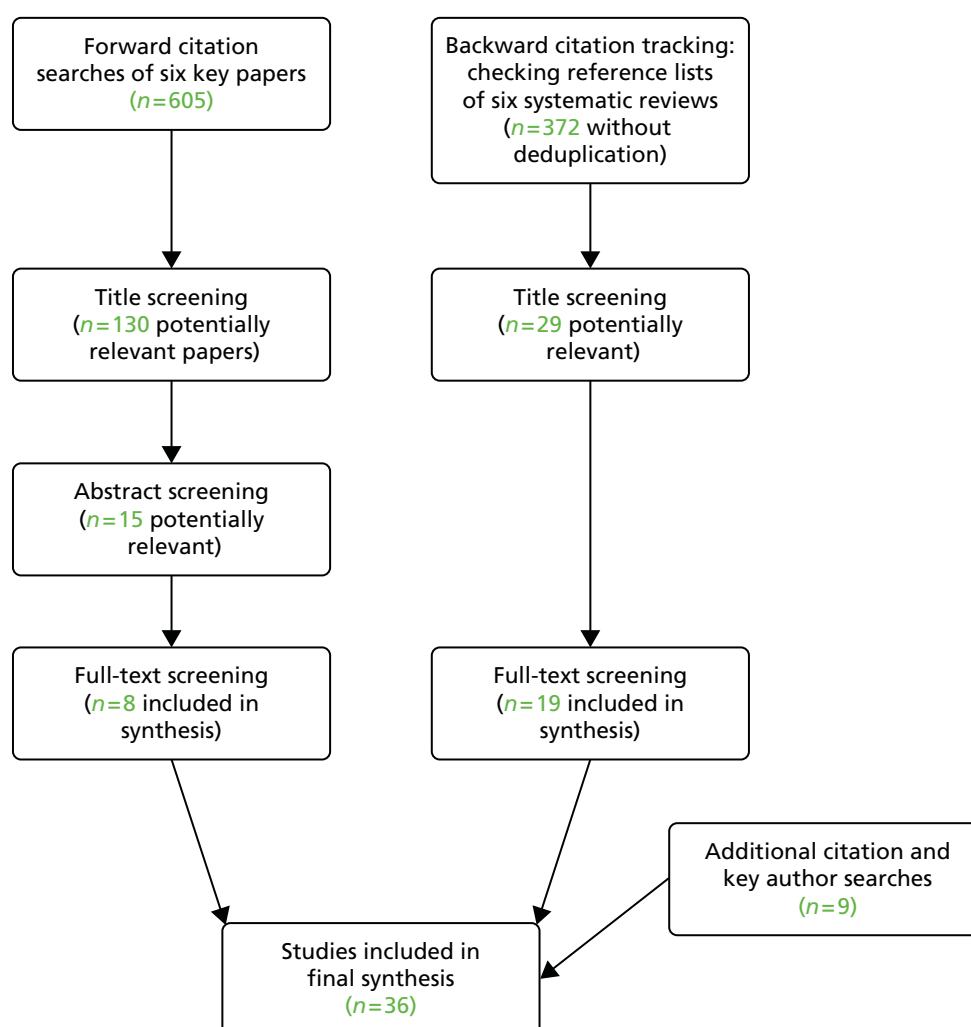


FIGURE 17 Process of study review and selection for the individual-level synthesis.

Together with our programme theories, we developed an initial logic model of the implementation chain from PROMs feedback to outcomes. To test and refine this model, we initially selected six papers^{63,65,81,242–244} and conducted a ‘mini’ or pilot’ synthesis on these papers. The papers were selected to represent different contexts, different points in the PROMs implementation chain and different types of PROMs. Papers were summarised using a data extraction template (see *Appendix 2*), and a cross- and within-analysis of the papers was carried out. In this pilot synthesis we sought to explore:

- How do patients experience PROMs completion; do they complete PROMs ‘honestly’?
- How do PROMs support or constrain patient–clinician communication?
- How do clinicians use PROMs in decision-making?
- How/why does the type of PROM make a difference?
- How/why does the condition or setting make a difference?

We discussed our initial findings with the project group and this informed the final focus of the review (described above). We focused on two key overall programme theories:

1. PROMs act as a tool to support patients to share or raise concerns with clinicians.
2. PROMs act as a tool to raise clinicians’ awareness of patients’ problems and promote discussion during the consultation.

We then identified studies that were relevant to the testing of these theories. The studies were summarised initially in a data extraction table, which outlined the aims, methodology, findings, authors’ interpretation of those findings and how the paper linked to our programme theories. This facilitated a comparison of the contexts within which and processes through which PROMs feedback did or did not lead to the anticipated outcomes. JG, EG and KG discussed our ongoing synthesis through regular Skype meetings and with feedback from the wider project group through a series of working papers and project group meetings.

We explored the first of these programme theories, about patients communicating their concerns, by assembling studies that explored patients’ experiences of completing standardised and individualised measures, and by comparing patients’ and clinicians’ experiences of PROMs completion in palliative care and oncology versus in the care of people with mental health problems in primary and secondary care. The findings are reported in *Chapter 8*. To explore the second programme theory, on clinician awareness and consultation discussion, we used oncology as a case study and examined outcome patterns across different RCTs, and attempted to explain these by examining surveys and qualitative and quantitative studies of clinician–patient interactions. The findings are reported in *Chapter 9*. Again, we wrote our synthesis up as a narrative account of each study and did so for a number of reasons. First, it enabled us to show how each study contributed to the theory-testing process and in effect ‘show our working out’ so that the reader can clearly see how we came to our conclusions. Second, it enabled us to incorporate an assessment of the study’s quality and highlight any caveats the reader needs to be aware of in the actual narrative of the synthesis, rather than as an assessment that remains separate to the synthesis findings. We provide a number of summary sections in each chapter to enable the reader to take stock of our findings.

Chapter summary

In this chapter we have described in detail the process through which we conducted a realist synthesis to explore in what circumstances and through what processes does the feedback of individual PROMs data improve patient care. We followed the RAMESES guidelines⁸³ to report our synthesis methods and have described:

- the rationale for our review
- the objectives and focus for our review

- how the boundaries of our review were defined and progressively focused, to explain how changes to the review process from the protocol⁸⁸ were made
- how we searched for and identified our candidate programme theories
- how we searched for and selected empirical evidence to test and refine our candidate programme theories
- how we carried out the appraisal and data extraction
- how we carried out the analysis and synthesis process.

Chapter 7 reports the findings of our theory elicitation process and *Chapters 8* and *9* report the findings of our evidence synthesis. A summary of the theories tested and studies included can be found in *Appendix 4*.

Chapter 7 Candidate programme theories of how the feedback of patient-reported outcome measures is intended to improve the care of individual patients

Patient-reported outcome measures in the care of individual patients: a brief programme history

The majority of the currently available PROMs were originally designed for use in research to ensure that the patient's perspective was integrated into assessments of the effectiveness and cost-effectiveness of care and treatment.² Their use in this context stemmed from the argument that clinical or biomedical measures of treatment impact did not capture outcomes that matter to patients; treatments may be deemed successful on the basis of biomedical criteria but may have little or even a detrimental impact on patient functioning.²⁴⁵ A classic example of this tension was the findings of the Diabetes Control and Complications Trial,²⁴⁶ which compared intensive therapy administered either with an external insulin pump or by three or more daily insulin injections together with frequent blood glucose monitoring with conventional therapy with one or two daily insulin injections for people with insulin-dependent diabetes. The trial showed that although intensive therapy improved metabolic control and reduced the incident of long-term complications, it also increased the incidence of hypoglycaemia in the short term. This demonstrates the trade-off between long-term and short-term outcomes in assessing the effectiveness of treatments for diabetes. It was assumed that there was a strong link between clinical end points and a patient's 'quality of life' but numerous studies have shown only a weak link between the two.²⁴⁵ As many treatments for chronic disease focus on improving not just the length of patients' lives but also the quality of their lives, a strong argument was made that clinical trials should also assess the impact of treatments on a patient's own perceptions of his or her health.²⁴⁷

These concerns led to a rise in prominence of the concept of 'health-related quality of life' (HRQoL) and the proliferation of instruments designed to measure it. Precise definitions of the concept of HRQoL were contested; some defined HRQoL as 'those parts of quality of life that directly relate to an individual's health' (p. 25),²⁴⁸ while others argued that it is was not possible to separate HRQoL from quality of life and criticised the concept for a lack of consensus regarding its definition.²⁴⁹ At the heart of these debates lay the challenge of attribution; some aspects of a patient's quality of life were not amenable to change through interventions focused at improving patients' health status and, as such, it was questioned whether broader measures of quality of life were useful markers of treatment success. Attempts to address this conundrum included the development of models to show the pathway through which changes to clinical variables impacted on symptoms, which in turn impacted on a patient's functional status, general health perceptions and, ultimately, overall quality of life.²⁵⁰

Instrument developers, while not ignoring these debates, to a large extent did not resolve these conceptual problems but instead focused on the task of developing instruments. Consequently, the last 30 years have seen an exponential rise in the number and type of such instruments designed to measure HRQoL.²⁵¹ Over time, these instruments sought to measure a whole range of constructs including, for example, HRQoL, symptoms, functioning and activities of daily living. Consequently, a broader categorisation of instruments emerged: patient-reported outcomes (PROs) in the USA and PROMs in the UK. These are defined as questionnaires that measure patients' perceptions of the impact of a condition and its treatment on their health.¹

Research efforts centred on testing the psychometric properties of specific instruments in different patient populations: for example, generic measures such as the Short Form Questionnaire-36 items (SF-36),²⁵² utility measures such as the EQ-5D²⁵³ and disease-specific measures such as the OHS²⁵⁴ for musculoskeletal

conditions, and the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30)²⁵⁵ and the Functional Assessment of Cancer Therapy – General (FACT-G)²⁵⁶ for cancer. Key psychometric properties included validity – the extent to which an instrument measures what it intends to measure; reliability – the extent to which an instrument is free from random error and produces consistent results either within observers (test–retest reliability) or between observers (intra-rater reliability); and responsiveness – the ability of an instrument to detect change over time. This last criterion was particularly important for instruments used in RCTs.^{165,257}

This burgeoning of research effort also saw the emergence of bodies such as the International Society for Quality of Life Research (ISOQOL) in 1994 and their associated journal *Quality of Life Research*, which focus on supporting the development of such instruments and holding conferences to support advancements in the science of measurement. Alongside the development of measures came work to establish appropriate methods, criteria and minimum standards for the psychometric properties for use in research settings.^{257–260} One example of these criteria, developed by ISOQOL, is reproduced in *Table 3*.

TABLE 3 The ISOQOL’s minimum standards for the use of PROs in patient-centred and comparative effectiveness research

Criteria	Minimum standard
Conceptual and measurement model	A PRO measure should have documentation defining and describing the concept(s) included and the intended population(s) for use. In addition, there should be documentation of how the concept(s) are organized into a measurement model, including evidence for the dimensionality of the measure, how items relate to each measurement concept and the relationship among concepts included in the PRO measure
Reliability	The reliability of the PRO measure should preferably be at or above 0.70 for group-level comparisons, but may be lower if appropriately justified. Reliability can be estimated using a variety of methods including internal consistency reliability, test–retest reliability or item response theory. Each method should be justified
Content validity	A PRO measure should have evidence supporting its content validity, including evidence that patients and experts consider the content of the PRO measure relevant and comprehensive for the concept, population, and aim of the measurement application. This includes documentation of (1) qualitative and/or quantitative methods used to solicit and confirm attributes (i.e. concepts measured by the items) of the PRO relevant to the measurement application; (2) the characteristics of the participants included in the evaluation (e.g. race/ethnicity, culture, age, gender, socio-economic status, literacy level) with an emphasis on similarities or differences with respect to the target population; and (3) justification for the recall period for the application
Construct validity	A PRO should have evidence supporting its construct validity, including documentation of empirical findings that support predefined hypotheses of the expected associations among measures similar or dissimilar to the measured PRO
Responsiveness	A PRO measure for use in longitudinal research study should have evidence of responsiveness, including empirical evidence of changes in scores consistent with predefined hypotheses regarding changes in the measured PRO in the target population for the research application
Interpretability of scores	A PRO measure should have documentation to support interpretation of scores, including what low and high scores represent for the measured concept
Translation of the PRO measure	A PRO measure translated to one or more languages should have documentation of the methods used to translate and evaluate the PRO measure in each language. Studies should at least include evidence from qualitative methods (e.g. cognitive testing) to evaluate translation
Patient and investigator burden	A PRO measure must not be overly burdensome for patients or investigators. The length of the PRO measure should be considered in the context of other PRO measures included in the assessment, the frequency of PRO data collection, and the characteristics of the study population. The literacy demand of the items in the PRO measure should usually be as 6th grade education level or lower (i.e. 12 years old or lower); however, it should be appropriately justified for the context of the proposed application

Reproduced with kind permission from Springer Science+Business Media: *Quality of Life Research*, ISOQOL Recommends Minimum Standards for Patient-Reported Outcome Measures Used in Patient-Centered Outcomes and Comparative Effectiveness Research, vol. 22, 2013, pp. 1889–90, Reeve B, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al., figure 5, © Springer International Publishing AG, Part of Springer Science+Business Media.²⁶⁰

An ongoing criticism of PROMs was the ordinal nature of many of the instruments, meaning that the gap between scores of 5 and 6 on a particular PROM was not necessarily the same as the gap between scores of 6 and 7 and, therefore, strictly speaking, they should not be analysed using parametric statistics. In addition, the requirement for instruments with robust psychometric properties had led to the production of instruments with many items that were onerous for patients to complete. Furthermore, PROMs varied in their appropriateness for patient groups with different levels of severity, and often suffered from floor and ceiling effects, which limited their responsiveness to change. In response to these problems, a new generation of instruments was developed based on item response theory or Rasch analysis. These instruments differed from traditional psychometric methods by testing how far items within a measure fitted the requirements for interval level measurement along a single dimension. Each item could then be plotted on a 'ruler', allowing a more precise ordering of items according to their level of 'difficulty' or severity.

Criticisms were also raised about the degree to which standardised HRQoL instruments adequately captured the patient's perspective.^{261,262} Many of the early measures were not developed in collaboration with patients and items were developed based on clinical perspectives of what was important to patients.²⁶³ Furthermore, the standardised nature of many existing PROMs assumed that all items were equally relevant to patients and there was little scope for patients to indicate how important each dimension was important to them. This gave rise to the development of a number of individualised measures, such as the Schedule for the Evaluation of Individual Quality of Life (SEIQoL),²⁶⁴ the Measure Yourself Medical Outcomes Profile,²⁶⁵ the Disease Repercussion Profile²⁶⁶ and the Patient Generated Index.²⁶⁷ These instruments all allow some flexibility for patients to select problems or domains that are particularly important to them and/or rate how important a domain is to them individually. Furthermore, a consensus guideline for the development of standardised PROMs specified that patients should be directly involved in the item generation process.²⁶⁸

Thus, in summary, there has been a sharp increase in the number of PROMs available to measure almost every aspect of a patient's health status, symptoms, functioning and HRQoL. Most of these have been developed for use in research rather than in routine clinical practice. There have also been significant developments in the methodologies underpinning their development and testing, and research has largely focused on the development and psychometric testing of these instruments. As a result of these endeavours described in previous paragraphs, a large number of different types of instruments now exist, summarised in *Table 4*. We now turn to consider their role in the care of individual patients in routine clinical practice.

Candidate programme theories underlying patient-reported outcome measures feedback in the care of individual patients

In most countries, the use of PROMs for individual patient care and their use at an aggregate level as an indicator of the quality of patient care for performance management purposes have developed separately but in parallel.²⁶⁹ The collection and feedback of PROMs in the care of individual patients has not formed an explicit part of government policy in the UK. For example, it was never intended that data collected as part of the English national PROMs programme would be routinely fed back at an individual level to clinicians to inform their care of individual patients, although it is possible for clinicians to request data for specific patients.¹¹⁹ Rather, it represents an intervention that clinicians and academics have proposed as 'a good idea' to improve care for patients and have set about evaluating its effectiveness in a series of randomised controlled clinical trials conducted over the last 40 years. This evidence has been amassed and synthesised in a large number of systematic reviews,^{24,41-54} discussed in *Chapter 1*. In addition, guidelines have been drawn up to assist practitioners with selecting and implementing PROMs in clinical practice.^{11,261} We focus on understanding the programme theories of how PROMs feedback in the care of individual patients is intended to work.

To do this, we examine the implicit hypotheses underlying how the intervention has been expected to work within RCTs evaluating the effectiveness of this intervention together with their role as envisaged in

TABLE 4 Different types of PROMs

Type of PROM	Description and examples
Generic	These instruments aim to measure health and functioning in the general population. Items aim to be relevant to people both with and without illness, and to people with any condition. They tend to measure a number of different dimensions, for example physical functioning, psychological functioning and social functioning, summarised as a number of different subscale scores. They are most useful for comparing the health and functioning of different patient populations. However, they tend to be less responsive to change over time in specific patient populations than disease-specific measures. Examples include the SF-36 and the Nottingham Health Profile
Disease-specific	These instruments aim to capture the specific ways in which a condition or its treatment impacts on patients' health and functioning. They also tend to have a number of different dimensions producing a number of subscale scores. They tend to be more responsive to change over time than generic instruments. Instruments vary in the degree to which they can be described as disease specific; many instruments were designed to capture the impact of a broad category of conditions (e.g. the OHS is often used to assess the impact of a broad range of musculoskeletal conditions)
Utility-based	These instruments were designed to combine quality and quantity of life into a single score of between 0 and 1 for use in health economic evaluations comparing different treatments. Respondents completing the measure indicate their current severity level on a number of dimensions of health; for example, the original version of the EQ-5D has five health dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) that are each scored at three levels of severity. The combination of the five dimensions and three levels of severity places participants in a particular health state; for example, someone with no problems walking about, some problems with washing or dressing, some problems performing daily activities, no pain or discomfort and who was moderately anxious or depressed would have a health state of 12212. Societal values for each state are calculated a priori in a separate exercise using tools such as the standard gamble and the time trade-off. These tools essentially ask individuals to indicate how much time in a specific health state they would be willing to trade for 1 year in full health. A sample of health states are valued in this way, which provides a score for each health state from 0 to 1. Valuations for the health states not directly measured are then estimated using methods such as conjoint analysis
IRT/Rasch	The instruments are developed through testing how far items within a measure fit the requirements for interval level measurement along a single dimension. Items that do not fit the model can be modified until they fit the model or discarded. Each item can then be plotted on a ruler, allowing a more precise ordering of items according to their level of 'difficulty' or severity. Examples include PROMIS
Individualised	These are measures that aim to move away from the standardised items that form the basis of generic, disease-specific, utility-based and IRT-based PROMs. Individualised measures enable patients to select issues that are of greatest importance to them, to rate how they feel about those issues and, in some cases, to determine the weighting that should be given to those issues. Examples include the SeiQoL-DW and MYMOP

IRT, item response theory; MYMOP, Measure Yourself Medical Outcome Questionnaire; SeiQoL-DW, Schedule for the Evaluation of Individual Quality of Life Direct Weight.

opinion articles, letters and reviews. It is important to note that we make no comment here on the veracity or validity of claims made within these programme theories, although we do highlight the debates and counter-theories that have arisen in response to them. Testing the extent to which these theories provide a useful explanation of how PROMs feedback works in practice will be addressed in the next chapters on evidence synthesis. A number of different applications of PROMs in the care of individual patients have been proposed.^{12,13,270–273} We have adapted and amalgamated these taxonomies and focus on the programme theories underlying the use of PROMs as:

- screening tools – to aid the detection of mental health and functional problems
- clinical monitoring tools – to monitor the impact of treatment on patient functioning and inform the clinical management of patient conditions
- personalised care planning and patient self-management – to facilitate patient involvement in care planning and decision making and support patients in self-managing chronic conditions.

It should be noted here that the distinction between these different ideas is somewhat blurred; for example, supporting patient involvement in decisions about their care can also support the clinical management of their condition. Therefore, we use this taxonomy with these caveats in mind.

Patient-reported outcome measures as screening tools

Early trials of PROMs feedback in the 1970s and 1980s focused on the use of PROMs as screening tools to aid the detection of mental health and functional problems in primary care and later in hospital outpatient departments.²⁷⁴⁻²⁷⁷ The majority of work in this area has focused on the value of PROMs in screening for depression. Screening for depression involves 'the use of self-administered questionnaires or small sets of questions to identify patients who may have depression but who are not already diagnosed or being treated for depression'.²⁷⁸ In these trials, patients who had not been diagnosed with depression were asked to complete a depression screening questionnaire, such as the General Health Questionnaire or the Zung Self-rating Depression Scale, and, for patients randomised to the 'intervention' arm, the score on the questionnaire was then fed back to their GP, along with information on the score, above or below which the patient might be considered 'at greater probability of' suffering from depression. The use of PROMs as screening tools was premised on the idea that GPs were underdiagnosing depression in their patients and, consequently, were not treating or referring onwards this population of patients. The underlying, implicit assumption was that GPs were not aware that their patients were suffering from depression, and providing them with a score which represented the patient's likelihood of having depression would increase GPs' ability to detect depression in their patients and, consequently, they would take appropriate action to manage the condition.

Between 2003 and 2013, the UK QOF financially rewarded GPs for using a standardised depression measures. In 2010, NICE published a guideline for the management and treatment of depression in adults that did not recommend routine screening for depression. Rather, this guideline advised GPs to be vigilant about the possibility that the patient may have depression, especially for patients with a previous history of the condition or for patients with a chronic condition, and recommended that GPs ask patients about symptoms of depression if they suspected patients may be at risk.

Patient-reported outcome measures as clinical monitoring tools

Here PROMs are envisaged as tool to support the clinical management of the patient. The underlying programme theory is that PROMs offer a more comprehensive assessment of patients' problems than clinical questioning alone, and that regular, ongoing feedback of PROMs to clinicians will enable clinicians to reflect on whether or not the treatment was working for this patient and, if not, to change the treatment accordingly. Long and Fairfield¹⁶² argued that monitoring whether or not the desired outcomes were being achieved for individual patients was an essential component of evidence-based medicine:

*At an individual patient level outcomes data are, or should be, used by clinicians to monitor how well a treatment plan is working and how the desired outcomes are being achieved, with a view to modifying treatment as appropriate.*¹⁶²

This approach has underpinned the use of PROMs to monitor patients' progress in psychotherapy, where it is known as 'patient-focused research'. For example, Lambert *et al.*²⁷⁹ explain that patient-focused research:

*... is aimed at monitoring an individual patients' progress over the course of therapy. This ... information can serve as valuable feedback to the practitioner ... who can make attendant treatment modifications in real time.*²⁷⁹

The feedback of patient progress in psychotherapy is assumed to raise clinicians' awareness of the gap between the patient's desired progress in therapy and their actual progress, creating cognitive dissonance and thus motivating clinicians to modify their approach to helping their clients to improve progress, in line with audit and feedback theories. Sapyta *et al.*²⁸⁰ drew on Bickman's²⁸¹ contextual feedback intervention theory to explain how the feedback of client progress in therapy would prompt clinicians to change their therapeutic practices with a client in order to improve the client's progress:

*for clinicians to engage in self-regulated change they need ... knowledge (feedback) about a discrepancy between that goal and the current status ... this contradiction between what one wants to accomplish and what one has actually accomplished creates dissonance, which is psychologically uncomfortable (Aronson, 1999). This dissonance is what motivates change.*²⁸⁰

Here, the assumption is that client progress need only be shared with the clinician and does not need to be discussed with the patient for changes in clinician behaviour to occur. It is assumed that the client's progress in therapy reflects the skill or practices of the clinician and they are the focus of the feedback.

Patient-reported outcome measures have also been envisaged as a tool to monitor treatment impact of other chronic conditions to enable clinicians to make changes to a patient's treatment. Here, PROMs are assumed to act like a test result, similar to biomedical indicators such as blood pressure or HbA_{1c}, and provide another piece of information on which clinicians can base their decision-making. For example, Wagner *et al.*²⁸² evaluated the feedback of the SF-36 to neurologists caring for people with epilepsy. Patients completed the measure prior to their consultation, and the patient's current and previous scores along with norms for the US population were placed in the patient's notes used by the neurologists during the consultation on at least two occasions. Wagner *et al.*²⁸² hypothesised that:

*Health status information would reveal a patient's decline or improvement in functioning and well-being and provide additional information to the clinician . . . [this] would help the clinician to uncover problems and monitor treatment response and . . . assist the clinician in the management of . . . the patient.*²⁸²

Similarly, Søreide and Søreide²⁸³ anticipated the following function of PROMs feedback in the care of patients with gastrointestinal cancer:

*. . . surgeons . . . can use the results of PROMs instruments to track patient's functional status and QoL [quality of life] changes through treatment . . . Obtaining formalized and 'objective' results (although based on patients 'subjective' report) might help surgeons . . . better communicate with patients during their treatment.*²⁸³

The second of these quotations suggests that PROMs offer a more systematic approach to the collection of information about patients functioning than clinicians' standard methods of history taking and treatment monitoring. Although not identical, this shares many of the ideas and assumptions underlying the use of disease templates and indicators, introduced as part of the QOF in primary care,²⁸⁴ which aim to provide a structure to the ways in which GPs monitor patients' lifestyle behaviours and prompt them to offer lifestyle advice to patients or to use structured needs assessments in social care and health visiting to ensure that certain topics are addressed with clients.^{241,285}

In summary, PROMs feedback is assumed to help clinicians monitor the impact of patients' treatment on their health, which in turn may lead to changes in treatment, referrals or further tests to explore the problem. It is thought that this in turn will lead to improved patient outcomes. Greenhalgh *et al.*⁶² developed a model to depict the implementation chain from PROMs feedback to improvements in patient outcomes, reproduced in *Figure 18*.

Patient-reported outcome measures as a tool to support personalised care planning and self-management

The idea of 'collaborative personalised care planning' underpins much of the recent policy and think-tank literature on the management of long-term conditions. Coulter *et al.*²⁸⁶ in a report for The King's Fund described this as follows:

Collaborative personalised care planning aims to ensure that individuals' values and concerns shape the way in which they are supported to live and self-manage their long term condition(s). Instead of focusing on a standard set of disease management processes, this approach encourages people with long term conditions to work with clinicians to determine their specific needs and express informed preferences for treatment, lifestyle change and self-management support. Then, using a decision coaching process, they agree goals and action plans for implementing them, as well as a timetable for reviewing progress.

p. 7. Reproduced with permission from The King's Fund²⁸⁶

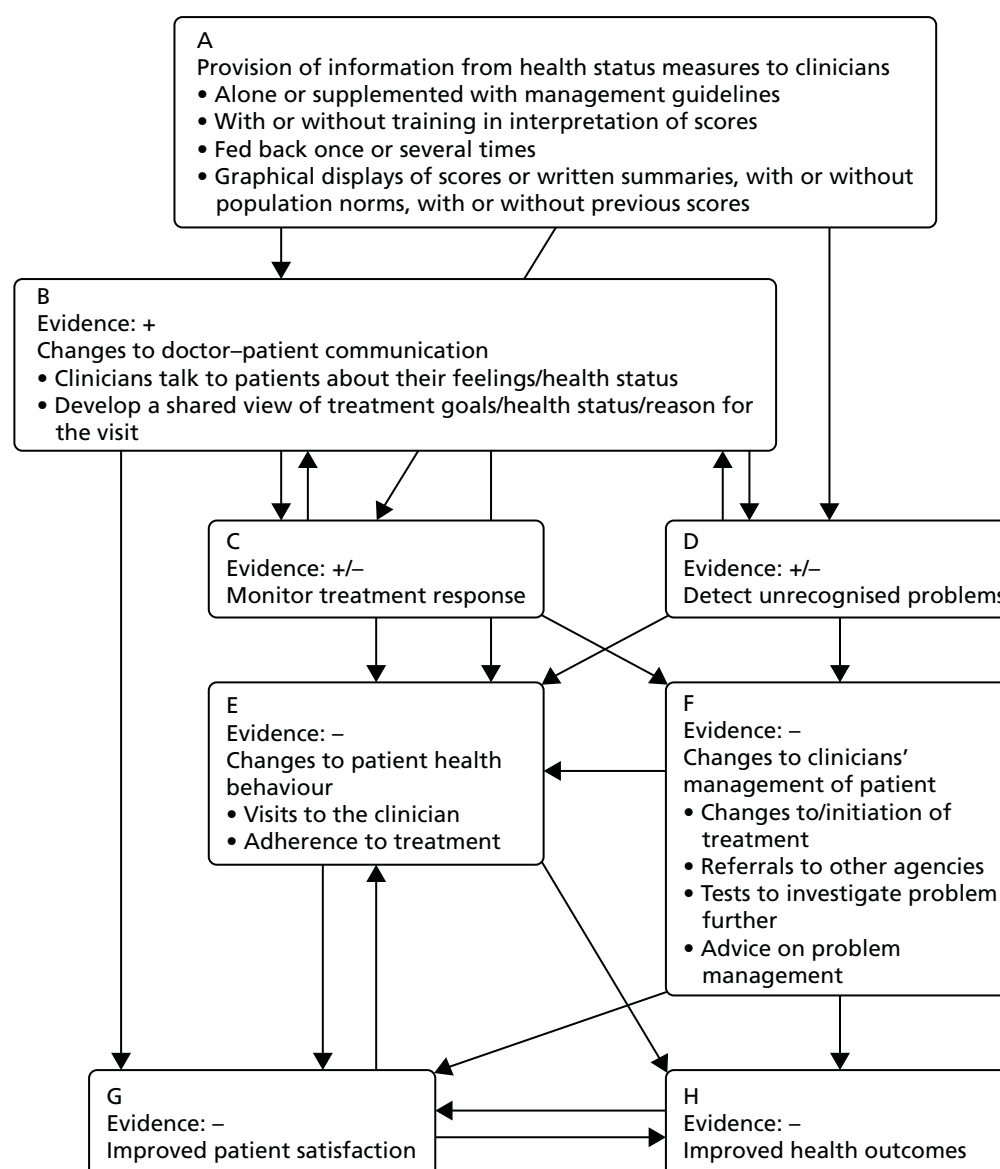


FIGURE 18 Greenhalgh *et al.*'s model of PROMs feedback in the care of individual patients. Reprinted from *Social Science and Medicine*, vol. 60, Greenhalgh J, Long AF, Flynn R, The use of patient reported outcome measures in clinical practice: lacking an impact or lacking a theory? pp. 833–43, © 2005,⁶² with permission from Elsevier.

Its model recognises that patients engage in most of their self-management activities during their day-to-day lives, away from the GP surgery or clinic. Therefore, the time they spend in consultation with clinicians presents an opportunity to better support a patient's own self-management. However, Coulter *et al.*²⁸⁶ also noted that:

It is acknowledged that having better conversations between clinicians and patients is not something that can be achieved without additional effort. Clinicians already have a structure for consultations 'hardwired' into their daily practice . . . The biggest change for clinicians involves recognising that the information about the lived experience and personal assets that the patient brings to the care planning process is as important as the clinical information in the medical record; processes also need to be in place to help the clinician identify and include the patient's contribution.

p. 7. Reproduced with permission from The King's Fund²⁸⁶

Thus, it acknowledges that clinicians may need to change the process of the consultation in order to recognise and incorporate the patient's perspective into the ways in which they discuss the condition with

the patient and make decisions. Coulter *et al.*²⁸⁶ presented a model of the consultation (*Figure 19*) that involves the integration of the clinician's expertise and assessment with the patient's knowledge of their condition to inform the management of the condition.

Patient-reported outcome measures have been envisaged as one tool that might support collaborative, patient-centred care and patient self-management.²³ For example, Santana and Feeny²⁸⁷ presented a model (*Figure 20*) of how the completion of PROMs may support the management of people with long-term conditions. The first step in their model is that PROMs may facilitate communication between patients and clinicians (but also between patients and their relatives and among different clinicians). They noted:

*Our framework theorizes that the completion of PROMs could affect communication among patient–clinician . . . by raising patient's awareness of his/her condition and facilitating the description of his/her symptoms to clinicians. Simultaneously, the provision of the information from PROMs to clinicians could trigger discussion of issues about which the patient is aware and concerned.*²⁸⁷

Similarly, Feldman-Stewart and Brundage²⁸⁸ envisaged that PROMs information can enable the patient to better describe their symptoms to clinicians in a language that clinicians can understand:

*. . . filing out the form . . . improves patients' skills at describing their symptoms, such as . . . identifying and classifying their symptoms . . . [this] resulted in the patient being more effective at conveying messages about his/her health state, which in turn were interpreted by the doctor and improved the doctors' beliefs about the patient's health states.*²⁸⁸

They also theorise that PROMs information may debunk the myth that if a patient does not mention a symptom during the consultation, it means that this symptom is not important to them:

*Providing PROs to physicians can help overcome a physician's belief that symptoms not mentioned by patients do not bother the patients.*²⁸⁸

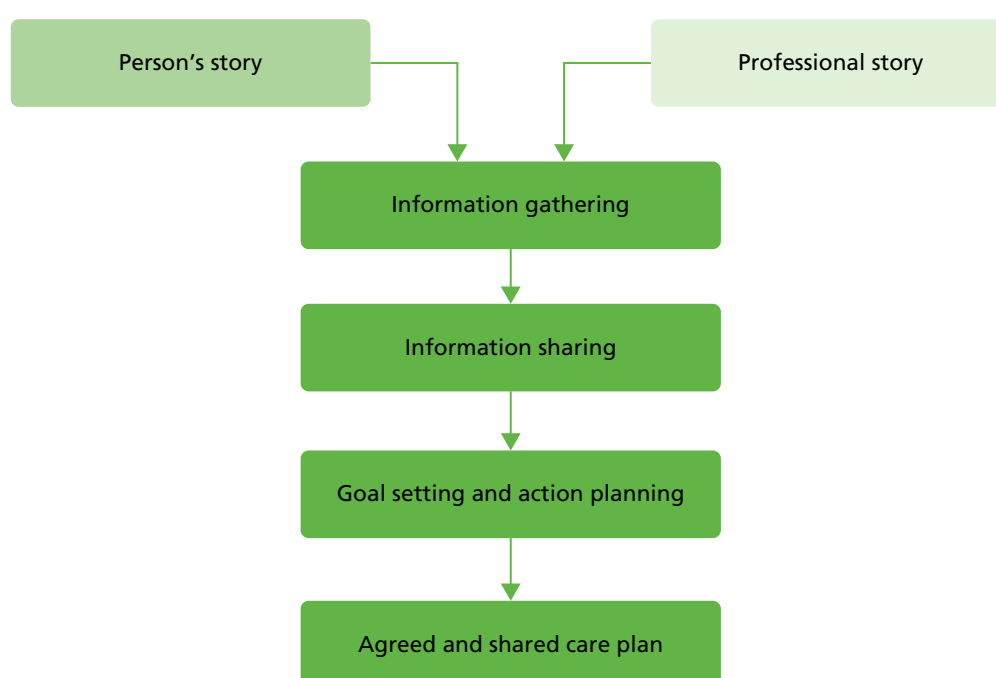


FIGURE 19 Model of the consultation in collaborative personalised care planning. Reproduced from Coulter *et al.*,²⁸⁶ with permission from The King's Fund.

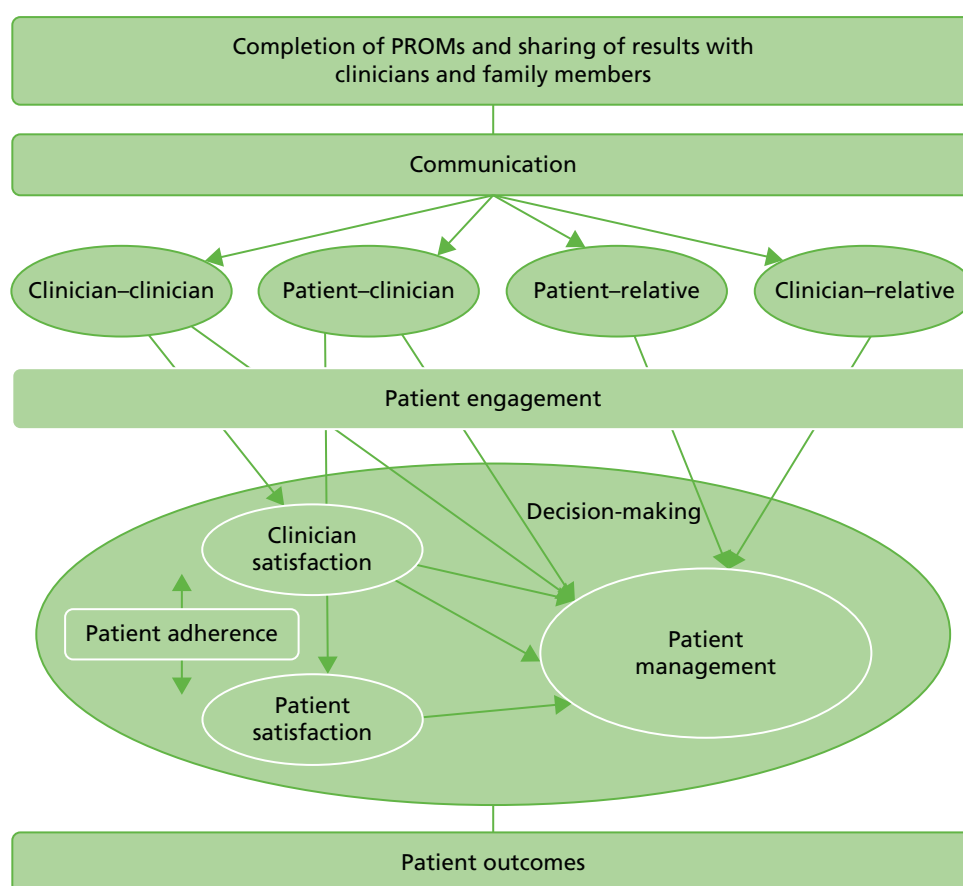


FIGURE 20 How PROMs completion can support the management of people with long-term conditions. Reproduced from *Quality of Life Research*, Framework to assess the effects of using patient-reported outcome measures in chronic care management, vol. 25, 2014, pp. 1505–13, Santana M, Feeny D, © Springer Science +Business Media Dordrecht 2013,²⁸⁷ with permission of Springer.

Snyder *et al.*¹¹ argued that providing PROMs information to clinicians may improve the efficiency of the consultation by highlighting issues that require addressing:

*The PRO results may be used to help prioritize the issues that require addressing in the clinic visit and promote efficiency.*¹¹

Santana and Feeny²⁸⁷ went on to envisage that improved communication stimulated by the use of PROMs may result in patients feeling more involved in their own care. Furthermore, clinicians may use PROMs to educate patients about their condition, which could further enhance patient engagement:

*We think that PROMs could improve communication between patient and clinicians involving patients in their own care. In situations in which clinicians use the PROMs data to discuss and educate patients, the use of PROMs data could have the potential to enhance patient engagement and activation.*²⁸⁷

Armed with increased knowledge about the patient's perspective, this, in turn can influence how clinicians make decisions about patient care management:

A potential effect of completing the PROMs may be that patients more frequently talk about the issues with the clinician and the clinician gains insight about patients' perspectives. Consequently, once clinicians recognize the issues as clinically important, they could initiate changes (ordering new tests, changing medications and dosages, and referring patients to other specialists) and monitor patients'

progression at the clinic visits as well as between visits. Potentially, this process could improve patient management. We assumed that the routine use of PROMs in chronic care management provides useful information to engage patients and their relatives more effectively and efficiently.

Reproduced from Quality of Life Research, Framework to assess the effects of using patient-reported outcome measures in chronic care management, vol. 25, 2014, pp. 1505–13, Santana M, Feeny D, © Springer Science+Business Media Dordrecht 2013, with permission of Springer²⁸⁷

However, the above description envisages the clinician as leading the decision-making on patient care. Santana and Feeny²⁸⁷ acknowledged that PROMs feedback could also support a shared model of decision-making:

PROMs data provide information about patient experiences and patient own preferences for health outcomes and the processes of treatment. Such information is not known by clinicians but is nonetheless important in choosing a specific treatment plan. The discussion between the clinician and the patient about the optimal treatment is of great importance given the availability of treatment options and the uncertainty of medical treatment outcomes.

Reproduced from Quality of Life Research, Framework to assess the effects of using patient-reported outcome measures in chronic care management, vol. 25, 2014, pp. 1505–13, Santana M, Feeny D, © Springer Science+Business Media Dordrecht 2013, with permission of Springer²⁸⁷

Thus, the overall programme theory is that PROMs may provide a vehicle through which the patients' concerns and perspectives about their condition and its management are integrated into the process of information sharing, goal setting and action planning to support both clinical management and the patient's own self-management of long-term conditions. In turn, clinicians and patients engage in a process of shared decision-making about care and treatment that reflects the patient's experiences and preferences.

Coulter *et al.*'s²⁸⁶ model implied that, for personalised care planning to occur, clinicians need to view the patient's perspective of their condition and its treatment as equally important as a biomedical perspective. It also suggests that a process of integrating the patient's perspective (offered through PROMs feedback and the patient's own descriptions) with clinicians' own clinical judgement and information from biomedical information (e.g. through tests and scans) is required. Furthermore, PROMs are envisaged as offering a systematic way of ensuring that patients' concerns are addressed.

Etkind *et al.*⁶⁰ presented a useful logic map (*Figure 21*) to summarise the steps through which PROMs feedback can improve the process and outcomes of care in the context of palliative care.

The models outlined above have varied in whether they conceptualise improvements in clinician–patient communication can have direct positive effects on the patient's well-being and/or whether this is mediated through improvements to the process of care, which in turn improve patient outcomes. Street *et al.*²⁸⁹ developed a theoretical model to explain how improvements in communication may have both direct benefits on psychological well-being and indirect benefits on health outcomes through changes to the process of care (*Figure 22*). For example, improvements to communication may have a direct influence on psychological well-being, as patients may gain therapeutic benefits when a clinician validates their perspective or expresses empathy towards them. However, improved communication can also have indirect benefits through involving patients in decision-making, which in turn increases the likelihood that the decisions made address patients' needs. They also explained that 'a clinicians clear explanations and expression of support could lead to greater patient trust and understanding of treatment options [proximal outcome] . . . this . . . may facilitate patient follow-through with a recommended therapy [intermediate outcome] which in turn improves . . . health outcome'.²⁸⁹

Finally, the electronic collection of PROMs data is hypothesised to facilitate the use of PROM in supporting patient-centred care and self-management. ePROs that are integrated into a patient's electronic health record can support patient self-care by enabling patients to monitor their own symptoms both during and

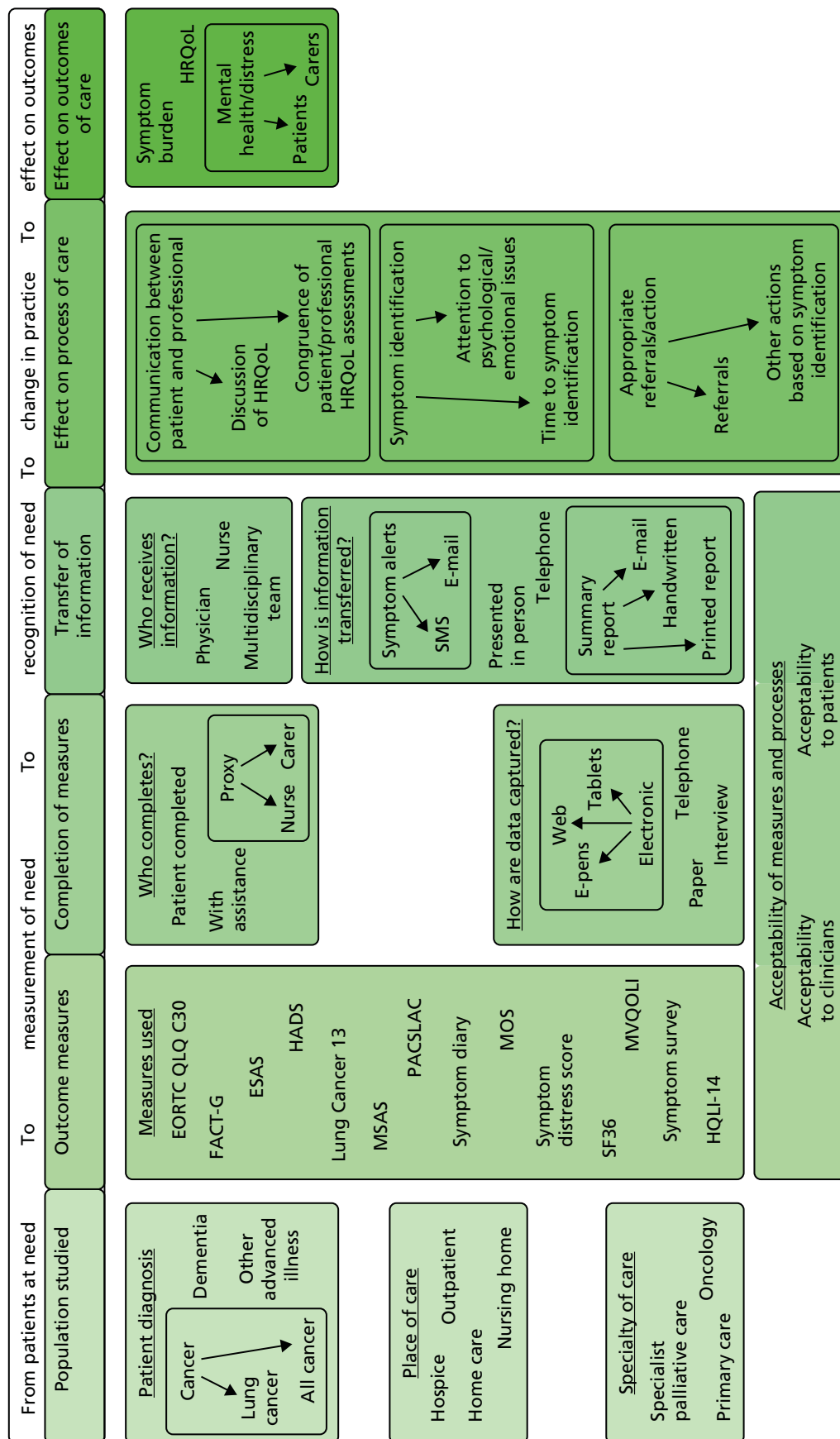


FIGURE 21 Etkind *et al.*'s⁶⁰ logic model of PROMs feedback in palliative care. Reproduced from Etkind *et al.* under the Creative Commons Attribution license (CC BY-NC-ND 4.0). ESAS, Edmonton Symptom Assessment Scale; HADS, Hospital Anxiety and Depression Scale; HQLI, Hospice Quality of Life Index; MOS, Medical Outcomes Study; MSAS, Memorial Symptom Assessment Scale; MVQOLI, Missoula-VITAS Quality of Life Index; PACSLAC, Pain Assessment Checklist for Seniors with Limited Ability to Communicate.

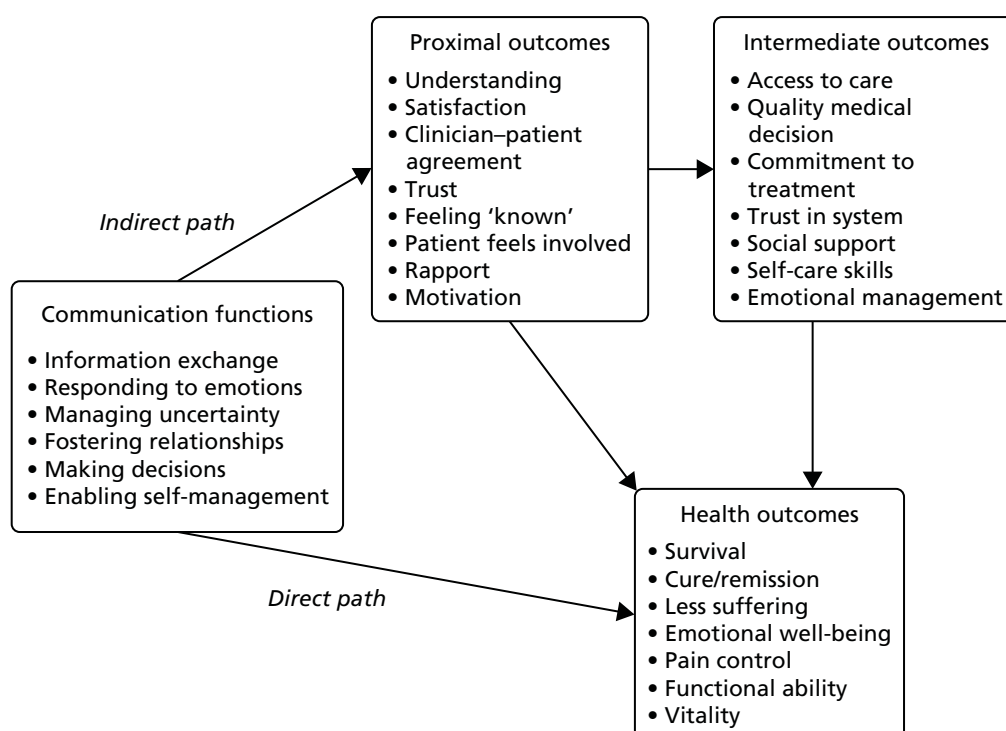


FIGURE 22 Street *et al.*'s²⁸⁹ model of the direct and indirect pathways from communication to outcomes. Reprinted from *Patient Education and Counselling*, vol. 74, Street RL, Makoul G, Arora NK, Epstein RM, How does communication heal? Pathways linking clinician-patient communication to health outcomes, pp. 295–301, © 2009,²⁸⁹ with permission from Elsevier.

after treatment and direct patients to self-care advice. For example, the Psychosocial Oncology and Clinical Practice Group at the University of Leeds have developed an online tool, eRAPID, to enable the remote collection of PROMs data that are linked to the patient's electronic health record.²⁹⁰ Warrington *et al.*²⁹¹ noted that:

*Support for self-management can be provided by offering specific semi-automated advice, based on PROMs scores linked to validated clinical algorithms. Such algorithms can direct survivors with persistent low-level problems to available self-help or community services.*²⁹¹

It is also thought to enable more efficient monitoring of patients at the end of treatment, to determine the level of follow-up and support they may require:

*... can be used for the initial assessment of cancer survivors at the end of their treatment to inform the individualised care plan, help with risk stratification and allocation to the appropriate care pathway (self-management, shared care or complex case management).*²⁹¹

Thus, it is theorised that patients may use ePROs to monitor and manage their own health independently from their contact with clinicians and to inform their decisions about whether or not and when to contact health professionals.

Counter programme theories: how the feedback of patient-reported outcome measures may not lead to the intended outcomes of improving patient care

We also identified a number of counter-theories that challenge the assumptions underlying the candidate programme theories outlined above. We outline these next.

Patient-reported outcome measures feedback is not practical or feasible

A number of counter-theories focus on the practicalities of collecting and feeding back PROMs during the clinical consultation. It is argued that a significant barrier to the collection and use of PROMs in the care of individual patients is the 'impracticality and burden of real-time administration and scoring of paper forms in the clinic'.²⁹² However, developments in IT now mean that patients can complete PROMs electronically through tablets and mobile phones and via the internet: so-called ePROs. It has been hypothesised that the electronic collection of PROMs data can facilitate their use in routine clinical practice by providing real-time data to both clinicians and patients. In support of this idea, Jensen *et al.*²⁹³ argued that:

*PRO use is . . . facilitated by . . . real time electronic platforms that collect, store and report PRO data to inform clinical care. Electronic (e-PRO) assessment systems allow efficient standardised assessments, decreased respondent burden . . . improved ease of use.*²⁹³

For example, Snyder *et al.*²⁹⁴ developed a website called Patient Viewpoint, which 'allows clinicians to assign PROs for the patient to complete, just as they would any laboratory test'.²⁹⁵ Snyder *et al.*²⁹⁴ indicated that the website was designed to:

*allow both patients and clinicians to track changes in status . . . so that the results of the PRO data can be evaluated in the context of the patient's other clinical information, the website links to an organization's EMR [electronic medical record].*²⁹⁴

Thus, the electronic collection of PROMs data is assumed to facilitate their collection and rapid feedback, which in turn is expected to enhance the use of data by clinicians. Furthermore, their integration with patients' electronic health records is theorised to enable clinicians to better integrate PROMs data with information from other biomedical or physiological tests into their decision-making.

Psychometric properties of current patient-reported outcome measures do not support their use in clinical practice

A number of counter-theories focused on the idea that the psychometric properties of current PROMs do not support their use in the care of individual patients. For example, critics of the use of depression screening questionnaires have argued that the majority of the original studies assessing the diagnostic accuracy of depression screening tools included patients who had already been diagnosed with depression, which may have exaggerated the accuracy of these tools.²⁹⁶ As such, it has been estimated that the number of patients currently not diagnosed with depression who would have been identified by screening tools 'may be less than half the number predicted by existing studies'.²⁹⁷ This may also increase the number of false positives (i.e. the number of people who are categorised by the screening tool as having depression but who do not have an underlying depressive disorder). These people may be prescribed unnecessary antidepressive medications, which are not without side effects. Thombs *et al.*²⁹⁷ also drew attention to the possibility of the 'nocebo effect', defined as 'the opposite to the placebo effect, whereby expectation of a negative outcome may lead to the worsening of a symptom'.²⁹⁸ Thombs *et al.*²⁹⁷ suggested that telling a patient that they have depression when they do not could trigger this effect and worsen outcomes for patients.

Goldberg,²⁹⁹ who developed one of the most commonly used depression screening questionnaires for research studies, the General Health Questionnaire, criticised 'the temptation for clinicians . . . to use screening questionnaires in too-simplistic a way, assuming that those with scores above some arbitrary threshold are psychiatric cases and those below are not'. To reduce the risk of false positives and false negatives in primary care, Goldberg advocated that, for patients who have a high score on the General Health Questionnaire, GPs should 'look at the questionnaire with the patient and ask additional probe questions suggested by particular symptoms' in order to ascertain whether the patient is likely to have a transient disorder or a more enduring problem that requires treatment.²⁹⁹ In other words, the screening tool should be used not as a definitive diagnostic tool but as a stimulus for further exploration by the

clinician. However, this requires time and effort on the part of the practitioner that they may not be able or willing to give. Furthermore, Gilbody *et al.*⁵¹ note that clinicians may 'intuitively recognise' that, in general practice, where the prevalence of depression is around 15%, only 50% of those with a positive screening result will actually have depression and, therefore, may be 'unwilling to act on positive test results'.⁵¹ Thus, rather than move to treat people who do not have depression, GPs may ignore the results of depression screening questionnaires completely as they distrust the results.

Similarly, it has also been argued that currently available measures were not sufficiently precise to permit their use in monitoring change in individual patients in clinical practice. For example, McHorney and Tarlov³⁰⁰ conducted a literature review to compare the psychometric properties of five commonly used generic PROMs and concluded that 'across all scales, reliability standards for individual assessment and monitoring were not satisfied'. However, in response, Hahn *et al.*³⁰¹ compared the reliability and measurement error of the same five generic PROMs with common clinical measures (e.g. blood glucose screening, forced expiratory volume measurements and systolic blood pressure monitoring). They concluded that 'by offering a juxtaposition of common medical measurements and their associated error with HRQoL measurement error, we note that HRQoL instruments are comparable with clinical data'.³⁰¹ They observed that HRQoL instruments are no more or less precise in monitoring individual patients than commonly used biomedical indicators, but continue to be perceived as less reliable, and are therefore rarely used by clinicians in their care of individual patients. Hahn *et al.*³⁰¹ argued that this is because clinicians are unfamiliar 'with the interpretation and potential utility of the data'; in other words, because clinicians rarely use such measures in their daily clinical work, they have not developed the so-called 'tacit knowledge' required to interpret and understand such data, as they have with commonly used biomedical indicators.

Patient-reported outcome measures may constrain rather than support the clinician–patient relationship

A number of counter-theories have focused on the idea that PROMs may constrain, rather than support, the clinician–patient relationship. As Ong *et al.*³⁰² argued, the clinician–patient consultation serves a number of functions, which include (1) creating a good interpersonal relationship, (2) exchanging information and (3) making treatment-related decisions. The underlying assumptions about how PROMs are intended to improve patient care anticipate improvements in all these functions. However, counter-theories suggest that the use of PROMs during the consultation may threaten the doctor–patient relationship because they do not sit easily with clinicians, who prefer to talk directly to the patient. Lohr and Zebrack³⁰³ explained that:

Practicing physicians tend to be both skeptical of and possibly irritated by pressures to use HRQOL instruments in daily practice. This skepticism pertains . . . to whether formulaic and standardized instruments provide any added value in eliciting information about their patients . . . the annoyance stems from the perceptions that researchers thought physicians were not doing enough or not doing right by their patients.

Quality of Life Research, *Using patient-reported outcomes in clinical practice: challenges and opportunities*, vol. 18, 2009, pp. 99–107, Lohr KL, Zebrack B, © Springer Science+Business Media B.V. 2008, with permission of Springer³⁰³

Similarly, Wright²⁶² noted that 'Clinicians usually do not rely on health status questionnaires in routine practice to judge the success of therapy' but instead 'feel more comfortable with actually asking them if they are better'. These quotations contain some implicit ideas about why clinicians do not use PROMs in the routine care of their patients. The first is that clinicians resent the implication that their current methods of history taking and talking to their patients are not sufficient to gather the appropriate and necessary information from patients to judge the success or otherwise of treatment. The second is that they question if PROMs, owing to their structured nature, are able to capture the individual concerns of patients. The authors also suggest that clinicians may feel uncomfortable using PROMs in their discussions with patients.

Other counter-theories assert that, rather than facilitating communication between clinicians and patients, PROMs feedback may damage the clinician–patient relationship and hinder communication. For example, Lohr and Zebrack³⁰³ expressed concern that clinicians may view PROMs as ‘short cuts to an appropriately complete and nuanced patient history’ and that patients may view PROMs as ‘offputting – a lesser substitute for true conversation and sharing’. They also warned that, rather than supporting the patient’s agenda, the use of PROMs in the care of individual patients could ‘detract from improving outcomes because they divert attention away from problems uppermost on the patient’s agenda and toward clinician-centred issues’.

Linked to this idea that standardised PROMs may constrain the clinician–patient relationship, others have argued that individualised measures may offer a solution to this problem. They note that standardised PROMs assume that all items are equally important to all patients and do not allow patients to indicate how important each item is to them and thus may not reflect the views of individual patients.³⁰⁴ It has been argued that individualised PROMs may be more appropriate than standardised PROMs for use in routine clinical practice, as they allow patients to nominate what is important to them and indicate how important that domain is to their HRQoL.^{305,306} Macduff³⁰⁷ argued that the process of completing individualised instruments such as the SEIQoL could provide the ‘therapeutic foundation’ for goal setting and developing the clinician–patient relationship, while the numerical data produced as a result of completion might be used as a measure of the effectiveness of clinician interventions. In other words, individualised measures can have value as a ‘conversation opener’ or vehicle for building the clinician–patient relationship, which is different from their use as a measure of the outcome of interventions. However, Macduff³⁰⁷ noted that the areas nominated by patients may change over time, which presents challenges in using individualised PROMs as indicators of the outcome of interventions. For example, it is not clear whether patients should be asked to rate the domains they originally nominated or whether it is acceptable to rate new areas identified by patients as important.

Other counter-theories question the broader idea that PROMs capture the patient’s perspective and thus offer a vehicle through which this can be more effectively communicated to clinicians and give primacy to the patient’s agenda during consultations. It is thus assumed that the balance of power between doctor and patient during the consultation will shift, with greater power being afforded to patients. However, Pilnick and Dingwall³⁰⁸ have questioned the basic premise of this argument. They note that, although interventions that attempt to increase the ‘patient centredness’ of consultations have changed the communication styles of clinicians and increased patient satisfaction, according to quantitative measures, observations of talk during consultations over many decades continue to show ‘the remarkable persistence of asymmetry’ in the doctor–patient relationship. They argue that this ‘asymmetry may have roots that are inaccessible to training programmes in talking practices’ and in fact ‘lies at the heart of the medical enterprise: it is founded in what doctors are there for’.³⁰⁸ In other words, the asymmetry is functional in the context of the wider social order; it enables doctors not just to care for patients but to maintain that social order by adjudicating on who has the right to be deemed ‘sick’. As such, it is not going to be changed by interventions that focus on changing communication practices during consultations.

Other counter-theories have focused on the challenges and costs that may be incurred by patients in the process of using ePROs to self-monitor and self-manage their own health. Lupton³⁰⁹ noted that it is assumed that through the practices of digital self-monitoring and self-care patients can have ‘control over one’s recalcitrant body and its ills’ and thus be empowered to take greater control of their health. However, she notes that digital self-monitoring may require patients to engage in self-monitoring at particular times of the day; thus, ‘empowerment’ becomes a ‘set of obligations’. She also observes that not all patients have the necessary economic or cultural capital to enact the ‘empowered consumer role’ that is envisaged by discourses of digital self-monitoring and self-care and that such patients may ‘find it difficult to challenge medical authority or simply may not wish to do so’.³⁰⁹ Finally, she argues that the very process of engaging in self-monitoring can be ‘too confronting, tiring or depressing for people who are chronically or acutely ill’.³⁰⁹

The challenges of using patient-reported outcome measures for multiple purposes

Another set of counter-theories focuses on the potentially unintended consequences of using PROMs for multiple purposes. Some of these theories consider how this may affect the behaviour of patients, while others consider how this might influence the behaviour of clinicians. Some have expressed concerns that when the results of PROMs are used by clinicians to determine access or continued use of treatments or therapies, patients may manipulate their answers to PROMs, which may misrepresent their true feelings but ensure that access to the desired treatment is maintained. For example, Lohr and Zebrack³⁰³ warned that:

*Will patients, deliberately or inadvertently, give misleading information on PRO instruments that might prompt unease, if not actual mistrust of patients by their doctors? . . . for instance, in using pain measures when patients are seeking narcotics or other prescription drugs for reasons other than pain per se . . .*³⁰³

Others have highlighted the challenges of using PROMs both for performance management purposes and in the care of individual patients.^{16,269} For example, Wolpert¹⁶ noted that PROMs data are often mandated for routine collection as measures of service quality without considering how such measures can be integrated with 'clinical conversations or clinical care'. She argued that the value of using PROMs data for audit purposes is often disconnected from the challenges faced by those tasked with implementing the measures on the ground, where they may undermine the clinical encounter. Wolpert explained that 'the standard questions may seem irrelevant to a given patient and can be experienced as a potential burden for clinicians and patients alike'.¹⁶ Thus, PROMs that may be useful as measures of service quality may not support clinicians in their care of individual patients; however, it is clinicians who are expected to collect these data.

An overall programme theory

In this chapter, we have presented a range of programme theories underlying how the feedback of individual-level PROMs data is intended to improve the care of individual patients in routine clinical practice. The data are envisaged as tools to:

- screen for patients' functional or mental health problems
- assist in the monitoring of treatment on patients' health, and inform clinical decision-making
- support patient-centred care and patient self-management.

We have also considered a range of counter-theories suggesting possible blockages to the implementation of PROMs feedback or explanations of why they may not work as intended. These theories formed the basis of the evidence synthesis, reported in the next chapters. To guide our evidence synthesis, we drew on the different logic models discussed above (e.g. Greenhalgh,⁶² Coulter,²⁸⁶ Santana²⁸⁷ and Etkind⁶⁰) to develop an overall implementation chain or logic model of the feedback of PROMs data within patient care (*Figure 23*).

This depicts the intermediate steps through which PROMs feedback may enable patients or clinicians to raise issues during the consultation, discuss the issues, act on the issues and subsequently improve patient outcomes. It also shows that PROMs feedback may also enable patients to monitor their own health independently of their interactions with clinicians and that clinicians may use PROMs to inform their care of patients independently of their interactions with patients. The figure is intended not to be exhaustive but to be illustrative of the process through which PROMs feedback is intended to inform and improve the care of individual patients in routine clinical practice. This model provided a framework for our synthesis, reported in *Chapters 8 and 9*.

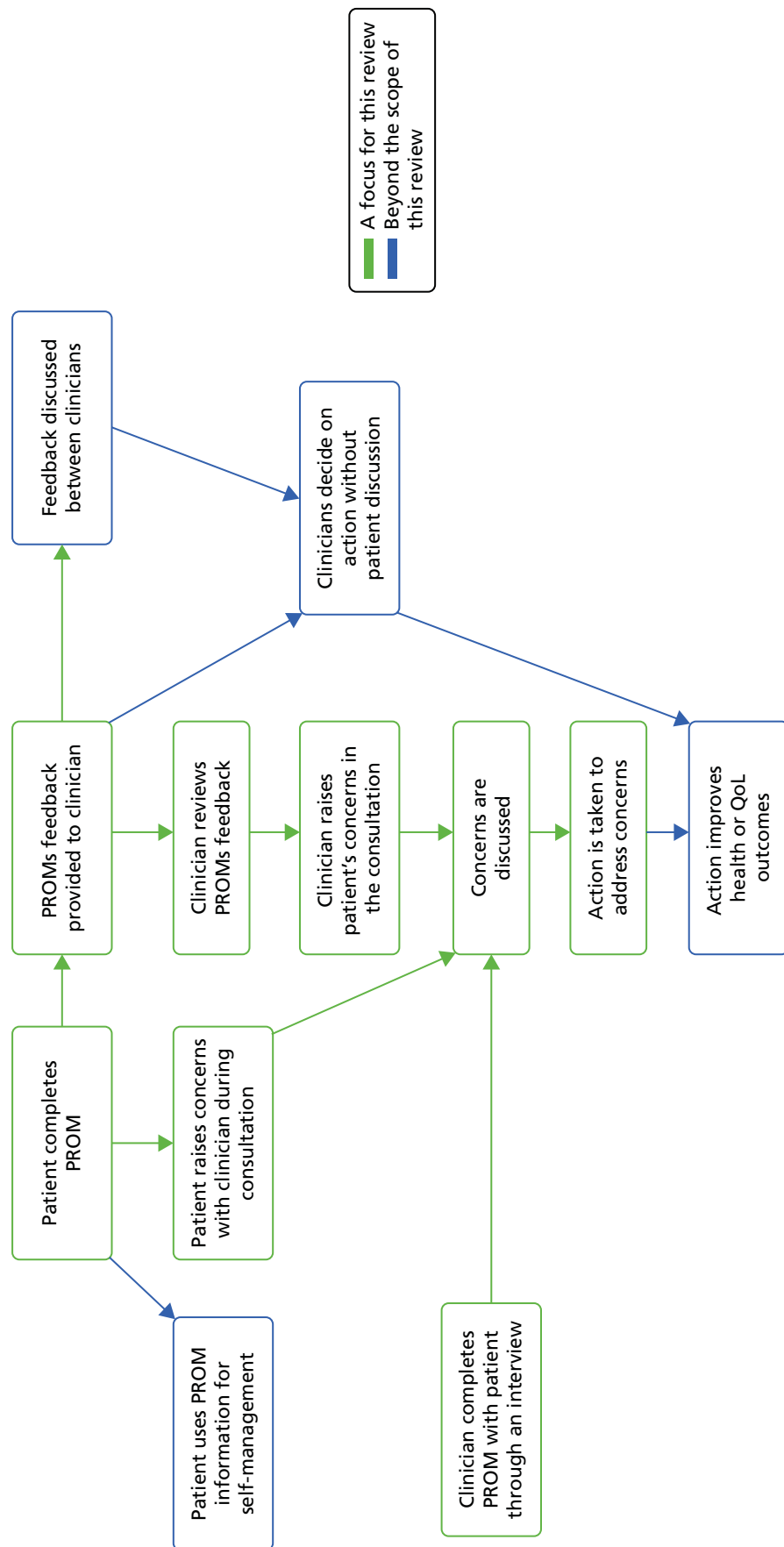


FIGURE 23 Patient-reported outcome measures feedback in the care of individual patients: a logic model.

Chapter 8 Patient-reported outcome measures as a tool to support patients in raising or sharing concerns with clinicians

Introduction

As we discussed in *Chapter 7*, the feedback of individual PROMs data to inform and support the care of individual patients can serve a number of different functions. PROMs can be used as tools for:

- screening – to aid the detection of mental health and functional problems
- clinical monitoring – to monitor the impact of treatment on patient functioning and inform the clinical management of patient conditions
- personalised care planning and patient self-management – to facilitate patient involvement in care planning and decision-making and support patients in self-managing long-term conditions.

As discussed in *Chapter 6*, following discussion with our stakeholders and patient group and within our project team, we agreed to focus our synthesis on understanding the mechanisms through which and contexts in which the feedback of PROMs data can help to support personalised care planning. We have used our logic model or ‘implementation chain’ of the pathways through which individual PROMs feedback may support personalised care planning, summarised in *Figure 23* in the previous chapter, to provide a structure for the review. This illustrates a number of pathways through which individual PROMs feedback may improve patient care. It also illustrates the proximal, intermediate and distal stages or outcomes of PROMs collection and feedback in the care of individual patients.

These pathways and stages provide a model for thinking about the way in which PROMs can support personalised care planning, the clinician–patient discussion and subsequent actions. In practice, these different stages overlap and they should not be understood as strictly sequential. For example, if PROM questionnaires are completed through an interview with a clinician, PROM completion and discussion may be simultaneous. Our synthesis is focused on understanding the mechanisms through which and contexts in which each stage in the pathway leads to the next and the potential blockages, obstacles and unintended consequences of PROMs feedback which may prevent or limit the achievement of its intended outcome of improving patient care. For example, the process of completing the PROM may help patients to articulate their concerns in the consultation, thereby supporting information sharing and discussion. We consider these processes and ways in which PROMs can be used in more detail below, and discuss the boundaries and scope of our review.

Patient-reported outcome measures as a tool to support clinical decision-making independently of the patient consultation

Patient-reported outcome measures may be used by clinicians to inform treatment decisions independently of patient consultations and without discussing the PROMs data with patients. This may involve clinicians using PROMs feedback to reflect on their own practice and change treatment accordingly, without discussing this with patients explicitly. This aligns with Bickman *et al.*'s²⁸¹ contextual feedback intervention theory, which assumes that the PROMs data will enable clinicians to see a gap between their own performance and the goal performance, in this instance that the patient is making the expected progress in therapy or that the patient's treatment is working as expected. It is assumed that, on being made aware of this gap, clinicians subsequently change their approach to therapy or the patient's treatment to improve the patient's outcome.

This programme theory underlies much of the research examining PROMs feedback within psychotherapy; for example, randomised trials have assessed the effects of providing feedback on patient progress to therapists, with a particular focus on highlighting patients who are not making the expected improvements.^{41,50,310} In these trials, therapists use the PROMs as part of their own review of whether or not therapy is working, without necessarily discussing the feedback as part of the patient consultation. Furthermore, standardised outcome measures have also been used by multidisciplinary teams to inform clinical decision-making.^{311,312} Again, these functions of PROMs occur 'backstage' to interactions with patients. We decided that these functions of PROMs were outside the scope of our review, as our focus was on how PROMs may support patients' involvement in the process of planning their care. However, we have drawn on some of the literature on patients' views of measures in these contexts, as the completion of PROMs by patients is common to all programme theories of PROMs.

Patient-reported outcome measures as a tool to promote patient self-management of long-term conditions

The process of self-monitoring using PROMs may support patients to self-manage long-term conditions, both in collaboration with and independently of their contact with clinicians. This is an important emerging area and a number of systems exist to capture PROMs electronically and, in some cases, link data to the patient's electronic health record.²⁹³ PROMs completion is often embedded in broader self-management support interventions, and a number of studies have examined the feasibility and acceptability of web-based PROMs completion^{313,314} and its impact on patient outcomes.³¹⁵⁻³¹⁸ The process of self-monitoring and its role in supporting self-management for people with long-term conditions has been considered in a number of existing systematic reviews.^{319,320} We have not focused on exploring whether or not and how patients use PROMs to self-manage long-term conditions.

Patient-reported outcome measures as a tool to facilitate patient self-reflection and to help patients raise or share issues with clinicians

The process of completing a PROM, by a patient either on their own or with a clinician, can make it more likely that the patient will raise issues of concern to them during the consultation. In *Chapter 7*, we outlined a number of mechanisms through which this is hypothesised to occur. PROMs completion may act as a simple reminder for patients to mention issues to clinicians. When completed independently or with a clinician, PROMs may enable patients to engage in a process of self-reflection and help them to prioritise which issues are important to them and to identify those that they wish to raise with clinicians. This process of self-reflection through PROMs completion may empower patients or give them 'permission' to raise issues with clinicians. It may also enable them to verbalise these concerns in a language that clinicians understand. It is assumed that the patient raises these issues with the clinician and the issues are then discussed.

Patient-reported outcome measures as tool for raising clinicians' awareness of patient concerns

A further pathway is that PROMs are a tool for raising clinicians' awareness of patient concerns. When a patient completes a PROM and the individual patient's PROM scores are fed back to the clinician prior to or during the consultation, it is hypothesised that this will alert the clinician to problems or issues that are of concern to the patient. These scores could be at a single point in time but also denote changes in PROMs scores over time, to enable the clinician to monitor the impact of treatment on the patient's health. It is assumed that this increased awareness will prompt the clinician to further explore and discuss any problems identified with the patient, and subsequently, to take action to address them, for example referral to additional services, advice, changing current treatment or other treatment options. This is then assumed to lead to improved patient outcomes.

Summary and structure of our review

Thus, our logic model has set out a number of different pathways through which the feedback of individual PROMs data is expected to improve patient care. It has also identified a number of proximal, intermediate and distal outcomes that are expected to occur as a result of this feedback. Our review will focus on two main pathways:

1. PROMs act as a tool to enable patients to raise or share concerns with clinicians.
2. PROMs act as a tool for raising clinicians' awareness of patient concerns.

Both of these pathways are expected to increase discussion between the patient and clinician about the patient's concerns and, in turn, the clinician is expected to take action to address these concerns. We will now explore the pathways through which these actions in turn lead to better outcomes.

For each of these pathways and each stage of the implementation chain, we considered the underlying assumptions and ideas about what needs to happen for the next stage of the chain to be realised. We ordered the findings from the empirical literature around these stages and compared what is expected to happen with what actually happens in practice.

In this chapter, we focus on testing and revising the main theory that PROMs act as a tool to enable patients to share or raise issues with clinicians. We also consider the initial stage of this pathway and review studies examining patients' views of and preferences for different types of PROM. In *Chapter 9*, we focus on testing and revising the theory that PROMs feedback acts as a tool for raising clinicians' awareness of patients' concerns. In *Chapter 9* we also consider the circumstances in which and processes through which PROMs feedback leads to the discussion of patients concerns and action to address them.

Candidate programme theories: patient-reported outcome measures as a tool to enable patients to raise or share issues with clinicians

First, we set out a number of candidate programme theories to be tested in this chapter. We draw on the process of theory elicitation undertaken in *Chapter 7* to identify some provisional theories that seek to explain how PROMs feedback is expected to work. It is important to note here that, initially, these theories represent the simple assumptions regarding the circumstances in which and processes through which PROMs feedback is expected to lead to intended outcomes. They are provisional, and the purpose of the synthesis is then to test these theories in relation to the literature and refine them to produce a more sophisticated understanding of how PROMs feedback works in practice. Our focus is on reviewing the evidence to test and refine the theory that the process of completing a PROM, by a patient either on their own or with a clinician, can make the patient more likely to raise or share issues of concern to them during the consultation.

The first stage in this pathway is that the patient completes the PROM. PROMs completion may act as a simple reminder for patients to mention issues to clinicians. When completed independently or with a clinician, PROMs may enable patients to engage in a process of self-reflection and help them to prioritise which issues are important to them and to identify those issues they wish to raise with clinicians. This process of self-reflection through PROMs completion may empower patients or give them 'permission' to raise issues with clinicians. It may also enable them to verbalise these concerns in a language the clinicians understand. Alternatively, clinicians may use the PROM to structure their discussion with patients and to explore the patients' views about their condition, its treatment and the impact of both on their health. In both pathways, it is assumed that the PROM offers a means of giving priority to the patient's agenda during the consultation, by providing a more comprehensive and systematic picture of their concerns. It is assumed that the patient then raises these issues with the clinician and these are discussed. We summarise these theories as follows.

- *Theory 10* Overall candidate programme theory: PROMs act as a tool to enable patients to raise or share issues with clinicians.

This may occur through a number of mechanisms:

- *Theory 10a* Self-reflection: patients reflect on their own situation, which then enables patients to identify and prioritise what is important to them and what they want to share or raise with clinicians.
- *Theory 10b* Reminder: the process of PROMs completion reminds patients to mention or raise issues with clinicians.
- *Theory 10c* Permission: the process of PROMs completion signals to the patient that someone is interested in how they are feeling and this gives them permission to share or raise issues with their clinician.

However, a number of theories reviewed in *Chapter 7* noted concerns that PROMs completion may hinder the development of the patient–clinician relationship or narrow the focus of the consultation.³⁰³ Patient–clinician interactions are spaces in which patients present their problems to clinicians and clinicians work with patients to make decisions about care and treatment, but they are also sites for relationship building and managing social interactions.^{302,308,321} Therefore, PROMs completion may work in unintended ways to produce unanticipated outcomes, depending on how they fit, or do not fit, with these existing processes. In summary:

- *Theory 11a* PROMs completion or review may constrain the development of the patient–clinician relationship.
- *Theory 11b* PROMs completion or review may hinder the flow of the consultation.

Patient-reported outcome measures feedback to support the care of individual patients is no less complex than the collection and feedback of aggregated PROMs and performance data. As we highlighted in *Chapters 4* and *5*, PROMs feedback is implemented in different ways, into contrasting organisations and services, alongside a range of accompanying programmes and incentives. These different contextual factors may sharpen or blunt the extent to which PROMs capture patients' views, support the process of self-reflection, and remind or give patients permission to raise or share their concerns with clinicians. Furthermore, the assortment of contextual factors do not exist in isolation from each other but merge, overlap and mutually influence each other to produce a complex range of different outcomes. In our synthesis we have recognised this; many of the papers we review represent different configurations of these contextual factors, and it is impossible to isolate one from the others. However, our focus has been on seeking to understand how and why different configurations of contextual factors lead to a complex pattern of outcomes, both intended and unintended. For the sake of clarity, we discuss each set of contextual factors in turn below.

The first set of contextual factors relates to the structure and format of the PROM. If the PROM is completed before the consultation, the measure needs to be easy to understand and complete by patients. If completed beforehand or used during the consultation, the PROM needs to adequately reflect the concerns of patients, in terms of both content and structure. A number of tensions arise here. In routine clinical practice, there is a need for measures to be brief in order to be feasible for use. However, they also need to be comprehensive enough to capture patients' perspectives.²⁶¹ Although the structured nature of standardised PROMs is hypothesised to enable a more systematic and comprehensive reflection of patients' concerns, a number of these measures have not been developed with the involvement of patients and, as such, they may not reflect their views.^{263,322} Furthermore, they may not be flexible enough to reflect the dynamic nature of patients' concerns or the different values that patients place on them.²⁶¹ For this reason, it has been hypothesised that individualised measures may be more appropriate for use in routine clinical practice.^{264–266} Thus, to summarise:

- *Theory 12* The structure and format of the PROM make a difference to whether or not the intended outcomes of PROM are achieved.

Particular dimensions that might be important are:

- *Theory 12a* Length versus comprehensiveness of the PROM.
- *Theory 12b* Standardised versus individualised PROMs.

The second set of contextual factors relates to the existing configuration of the patient–clinician relationship and consultation. Whether PROMs completion or review of findings by clinicians fits into the flow and function of patient–clinician interaction is likely to depend on the nature of the existing relationship between the patient and clinician, the point at which the PROM is completed in relation to the development of this relationship, and the patient’s and clinician’s own preferences for the ways in which relationships are built. Thus, to summarise:

- *Theory 13* Whether the completion and review of PROMs data supports or constrains the flow and function of patient–clinician interaction depends on:
 - theory 13a: the existing relationship between patients and clinicians
 - theory 13b: the point in the relationship-building process when the PROM is completed or reviewed
 - theory 13c: patients’ and clinicians’ existing preferences for relationship building.

The third set of contextual factors relates to the purpose of PROMs collection and the ways in which PROMs data will be used. In their systematic review, Boyce *et al.*⁵⁸ found that clinicians expressed concerns about the lack of clarity regarding whether PROMs data were intended for use to inform clinical care or to monitor the quality of the service. For example, in both the UK and New Zealand, the routine collection of PROMs in mental health services has been advocated as a means of supporting the care of individual patients and, at the same time, monitoring the quality of services. In the UK, the routine collection outcome data are mandated as part of the Mental Health and Learning Disability Minimum Dataset. Similarly, the use of standardised depression questionnaires for screening and monitoring people with depression in primary care was incentivised under the UK QOF between 2003 and 2013, and the results were used as indicators of the quality of care in general practice. This has raised concerns that attempting to achieve two different functions may undermine the successful use of PROMs data for either purpose.¹⁶

- *Theory 14* Whether patient-reported outcome measures data collection is incentivised or used for other purposes makes a difference to the achievement of its intended outcomes.

Evidence review: theory testing and refinement

In our synthesis, we do not review each of the above provisional theories separately. This is because contextual factors exist in configuration and shape the mechanisms through which PROMs completion and feedback works in multiple and complex ways. Furthermore, the studies we review here reflect these multiple context configurations and speak to a number of different theories at the same time. Rather, we test and refine our theories through exposure to studies which have explored different elements of the PROMs completion process in different settings. We start by reviewing studies which have examined patients’ views and preferences for different PROMs through experimental or comparative studies, to test theories about whether patients prefer long or shorter measures and prefer standardised or individualised PROMs. We then consider a range of studies that have examined patient and clinician experiences of PROMs completion in two settings that have been the focus of much work in PROMs: mental health and oncology/palliative care. These two settings present different configurations of contextual factors highlighted above, which enable us to understand how and why they support or constrain patients in sharing or raising issues with clinicians.

In mental health settings, we examine studies exploring how GPs have used standardised PROMs to detect and manage depression. In general practice, patients usually already have an existing relationship with their GP, and GPs must mobilise this relationship when patients present with suspected signs of depression in order to discuss referral and management options with them. In England, the use of these instruments was incentivised under the QOF to reward GPs for delivering high-quality care between 2003 and 2013. We also examine the use of PROMs with adult and child mental health services. Here, newly referred patients and the clinicians caring for them establish a relationship over time, which is also seen as instrumental in achieving the aims of therapy. In this context, the routine collection of outcome measures has been mandated as part of government policy in a number of countries (e.g. England, New Zealand and Australia), but has not been incentivised with monetary rewards and sanctions.

We also explore the use of PROMs in palliative care settings. The goal of palliative care is to maximise quality of life for those patients with incurable disease by controlling symptoms and addressing patients' psychological and spiritual concerns, suggesting that a systematic and comprehensive understanding of patients' quality of life is vitally important.³²³ For this, reason, the routine collection of PROMs in palliative care has been advocated and guidance produced on the selection and implementation of measures.³²⁴ However, the routine collection of PROMs in palliative care has not been part of formal government policy, nor has it been incentivised with monetary rewards and sanctions. There is no government-mandated, formal initiative to collect aggregated PROMs data as an indicator of service quality in palliative care. The contrast between mental health and oncology thus provides a useful means of comparing use of PROMs in services where the collection of PROMs data has been incentivised or mandated in order to both support individual care and measure service quality (such as care of people with mental health), with the use of PROMs in services where it has not (such as oncology or palliative care).

Finally, we note here that many of the studies included in our synthesis are small-scale qualitative studies. When such studies are considered in isolation, caveats regarding the limited generalisability of each study apply. However, realist synthesis involves a process of consolidation, in which a number of different studies are brought together to test and refine a theory. The function of this process is to develop and enhance our explanation of the circumstances in which and processes through which interventions work. In such cases, confidence in the validity of an explanation increases when findings from different studies are replicated. Furthermore, in realist synthesis, data and evidence comprise not just the study findings but also the authors' own interpretations and explanations of those findings. Therefore, trust in the validity of an explanation is also buttressed when not only the findings themselves are replicated, but the authors' own explanations for these findings also recur.

Patients' views on the structure and format of patient-reported outcome measures

Traditionally, quantitative psychometric methodologies have been the predominant means of establishing the validity of PROMs. These assess the extent to which a new PROM correlates with existing PROMs thought to measure a similar concept or is able to distinguish between different groups of patients who, a priori, are expected to have different scores on the measure. In addition, the importance of involving patients not only in the item generation process, but also in assessments of face and content validity of resulting PROMs (i.e. whether or not items look to address relevant issues) and in checking that items are easy to understand and complete, has been recognised.^{263,268} In this section, we review a number of studies that provide a comparative assessment of patients' views of different PROMs. These studies allow us to test the theories of whether patients prefer shorter, less comprehensive measures or longer, more comprehensive measures. They also allow us to test the theory of whether patients prefer standardised or individualised PROMs.

Snyder et al.³²⁵

This study compared patients' views of three standardised PROMs designed for use with people with cancer using a RCT design. The three PROMs in question were the EORTC QLQ-C30, six domains from PROMIS, and the Supportive Care Needs Survey-Short Form (SCNS-SF-34). The EORTC QLQ-C30 is a disease-specific PROM originally developed for use in trials to assess the impact of cancer treatment on

patients' quality of life, but it has also been used in trials examining the feedback of individual PROMs data in routine clinical practice.^{326,327} It has 30 items that cover five functional domains (physical, role, emotional, social and cognitive) and eight symptom scales (fatigue, pain, nausea and vomiting, dyspnoea, insomnia, appetite loss, constipation and diarrhoea). PROMIS is a collection of short-form scales, item banks and computer-adaptive tests designed to be relevant to people across a range of conditions, including cancer. In this study, the fixed-item scales for anxiety, depression, fatigue, pain impact, physical functioning and satisfaction with social roles were used. The SCNS-SF-34 was designed to assess the care needs of people with cancer. It has 34 items covering five need domains (psychological, health systems and information, physical and daily living, patient care and support and sexual).

Those recruited for the study were patients of 12 clinicians involved in the care of people with cancer (medical oncologists, radiation oncologists and nurse practitioners) in the USA. Of the 301 eligible patients, 224 (74%) agreed to participate and were randomised to complete the EORTC QLQ-C30, the SCNS-SF-34 or PROMIS scales before every clinic visit during their treatment (weekly for radiation oncology, every 2–4 weeks for medical oncology patients) via the online Patient Viewpoint web tool. The PROMs were automatically scored by the Patient Viewpoint tool and for phase 1 of the study were printed and fed back to clinicians prior to the consultation. In phase 2 they were made available via a printed version, online via the Patient Viewpoint website or in the patient's electronic record. At the end of their treatment, patients were asked to rate their views of the PROM using 11 Likert scales in the form of the patient feedback form. The items in this questionnaire assessed the extent to which patients thought that:

- the PROM was easy to complete
- the PROM was easy to understand
- the PROM was useful to complete
- the PROM helped them to remember when they met the doctor
- the doctor used information for their care
- the PROM improved quality of care
- the PROM improved communication with doctor
- they felt more in control of their care
- they would recommend the PROM to others
- they would want to continue responding in the future.

The primary outcome measure was the percentage of patients who answered strongly agree or agree to all 11 rating items on the patient feedback form. In addition, an exit interview was conducted with all patients to further explore whether or not scores were discussed and addressed.

Of the 224 patients who enrolled to the trial, 181 completed their allocated PROM on one or more occasions and completed one or more items on the feedback form. For these patients, there was a statistically significant difference in the primary outcome measure for the three arms of the trial, with the EORTC QLQ-C30 being rated most favourably (74% strongly agreed/agreed to all items), followed by PROMIS (61% patients strongly agreed/agreed) and then the SCNS-SF-34 (52% patients strongly agreed/agreed). The individual item ratings on the feedback form varied both within and across each arm. All items on the feedback form were relatively highly rated for all PROMs (the lowest was 67% agreeing or strongly agreeing that the SCNS-SF-34 improved the quality of their care). However, there was also a consistent pattern of responses to the feedback form items, whereby, for all PROMs, a higher percentage of patients reported that the PROM was easy to understand and complete and helped them to remember when they met the doctor, than felt that the doctor had used the information for their care or that the PROM improved their quality of care or communication with the doctor. Care must be taken in making inferences from these patterns, as testing for variation within the items across PROMs was not part of the planned analysis of the trial. However, the data suggest a hypothesis that may be tested elsewhere: that the process of PROMs completion itself may have more value to patients than the feedback has to clinicians.

The exit interviews provided some explanations as to why the SCNS-SF-34 was rated less highly by patients. The authors report that 'several' patients felt that a number of the items were unlikely to change over time and were therefore 'not suited to repeat administration'. This may have been because the SCNS-SF-34 used a 1-month recall period but patients were completing the PROM weekly. Thus, the authors concluded that 'the measure does matter to patients' and PROMs vary in their perceived usefulness to patients. However, patients did not directly compare the three measures, and the findings might have been different had all patients completed all PROMs. In addition, this study is based on self-reports of patients' satisfaction with the PROMs under study, which are often overwhelmingly positive. In terms of the theory under test, the study suggests that 'not all PROMs are equal', and that they vary in their ease of completion and usefulness to patients. It also tentatively suggests a hypothesis that the process of completing the PROM is perceived as having more value to patients in terms of helping them remember things to discuss things with their doctor than in relation to the impact on doctor–patient communication.

Nilsson *et al.*³²⁸

This study provides a comparative analysis of patients' views of two commonly used generic PROMs: the SF-36 and the EQ-5D. The SF-36 has 36 items that cover eight subscales (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role emotional and mental health). Items vary in their response options; some items have three-level response options and others have five-level response options. The EQ-5D has five items covering mobility, self-care, usual activities, pain/discomfort and anxiety/depression. In the version assessed by this study, each item has a three-level response option. As such, this study provides a test of whether shorter, more general PROMs with less complex response options are preferred over longer, more comprehensive PROMs with more complex response options, or vice versa.

A total of 573 patients with a range of long-term conditions were recruited via health professionals from hospitals across Sweden. All patients received an intervention of some kind; these included patient education, surgical and pharmacological treatments. Patients were asked to complete the EQ-5D, followed by the SF-36, before and after their intervention. They were then asked to evaluate each instrument using a questionnaire. Two items asked patients to rate how easy the measure was to understand and respond on a 5-point scale from 'very easy' to 'very hard'. One item asked patients to rate how well they felt each questionnaire described their health in a comprehensive way, from 'very good' to 'very bad'. Patients were then asked if they felt that the routine collection of health outcome assessment would be valuable in the future (yes/no) and, if so, which questionnaire they would prefer (SF-36, EQ-5D or no preference). Patients were also given space on the form to expand on their answers.

Of the 573 patients recruited, 463 (80%) completed both measures and an evaluation form. *Table 5* summarises the findings of the study.

This table shows that patients rated the EQ-5D as significantly easier than the SF-36 to understand and respond to. There were no differences in patients' ratings in terms of the extent to which they perceived each measure to describe their health in a comprehensive way. The majority of patients expressed no preference for either measure to be routinely collected in the future. Of those who did express a preference, significantly more patients preferred the SF-36 than the EQ-5D.

The open-ended responses provide some explanation for these findings. Twenty-seven responses referred to the value of outcome assessment in general and 62 responses explained their preference (or lack of) for one instrument or the other. Patients commented on the positive aspects of outcome assessment, including that it 'saves time at consultations', shows that 'someone cares!' and may give 'a better insight into the patient's experience'. Patients also commented that completing the questionnaires 'forces you to reflect on your situation'. 'Negative' comments included scepticism that it was possible to measure health using questionnaires, reflected in the comments 'how can you measure health on paper?' and 'mental or physical health is difficult to describe in text'.

TABLE 5 Patient ratings for the SF-36 and EQ-5D

SF-36	EQ-5D	p-value	
<i>Ease of understanding: percentage of patients rating questionnaire very easy/easy to understand</i>			
70	75	0.005	
<i>Ease of responding: percentage of patients rating questionnaires as very easy/easy to respond to</i>			
54	60	0.001	
<i>Ability to describe health: percentage of patients rating very good/good ability to describe health in a comprehensive way</i>			
68	68	0.88	
Preferred SF-36, % (n)	Preferred EQ-5D, % (n)	No preference	p-value (SF-36 vs. EQ-5D)
<i>Preference for instrument in routine care</i>			
25 (52)	8 (16)	68 (142)	< 0.001
Quality of Life Research, Respondent satisfaction regarding SF-36 and EQ-5D, and patient's perspectives concerning health outcome assessment within routine care, vol. 16, 2007, pp. 1647–54, Nilsson E, Wenemark M, Bendtsen P, Kristenson M, © Springer Science+Business Media B.V. 2007, with permission of Springer. ³²⁸			

Of those patients who expressed no preference for one measure or the other, free-text responses reflected that the questionnaires both seemed similar, were equally easy to respond to or answer, or were both equally incomplete. Those who preferred the SF-36 did so because they felt that the content of questions or the nature of the response options provided a more comprehensive assessment of their health, as illustrated in the quotations 'more detailed questions which give a better picture of how you feel' and 'more response options to choose from, which makes it easier to describe your condition'. Those who preferred the EQ-5D did so because they found it easier to respond to – 'easier to understand' – or less confusing than the SF-36: 'SF-36 is awkward to answer'. Some patients also identified problems with the EQ-5D, despite indicating a preference for it; these included that they were 'too healthy for the questions' or that the response options were too coarse: 'I would like to answer in between the response alternatives'.

This study is based on patient responses to a relatively simplistic questionnaire to explore their views of two PROMs. All patients completed the PROMs in the same order and their preferences might have been subject to 'order effects'. However, the study provides a useful test to explore whether patients prefer shorter, less comprehensive PROMs or longer, more comprehensive PROMs. Overall, quantitative data collected in this study found no differences in the extent to which patients felt that each questionnaire adequately described their health, but they felt the EQ-5D was easier to understand and respond to. Although most patients did not express a preference, of those who did, more patients would prefer the SF-36 to be routinely collected than the EQ-5D. The open-ended responses suggested that patients who preferred the EQ-5D did so because it was easier to understand, while the SF-36 was chosen because it was perceived to provide a better reflection of patients' health. In terms of the theory under test, this study highlights the tension between simplicity and complexity and between brevity and comprehensiveness in the structure and format of PROMs. Overall, it suggests that patients with long-term conditions are willing to tolerate a questionnaire that is longer and more complex to complete if it provides a more comprehensive assessment of their health.

Neudert *et al.*³²⁹

This study assessed patients' perceptions of the validity and distress caused by completion of three different PROMs: the Sickness Impact Profile (SIP), the SF-36²⁵² and the Schedule for the Evaluation of Individual Quality of Life Direct Weight (SEIQoL-DW). The SF-36 has been described previously. The SIP is a standardised generic PROM and consists of 136 yes/no items covering 12 domains. The SEIQoL-DW is an individualised PROM developed for use by patients with a range of conditions.³³⁰ During completion,

patients are invited to choose five life areas they consider to be the most important determinant of their quality of life. They are then asked to rate how satisfied they are with each area on a visual analogue scale. They are then asked to indicate the relative importance of each area using a weighting instrument, such that a pie chart is produced containing five segments, each representing a chosen area, and where the size of each segment represents the relative importance to patients. As such, this study compared patients' preferences for standardised generic PROMs versus an individualised PROM. It therefore allows us to test the theory that patients consider individualised measures to have greater face validity than standardised PROMs.

Patients with amyotrophic lateral sclerosis at one neurological centre in Germany were invited to participate in the study; most of them were enrolled on a drug trial. All patients completed the SIP as part of the drug trial and were randomly allocated to also complete either the SF-36 or the SEIQoL-DW. Patients completed the SIP and SF-36 independently or with help from one of the authors, whereas the SEIQoL-DW was administered through an interview. Patients completed the measures at least three times at 2-month intervals. Patients were asked to rate their perceived validity and the degree of emotional distress involved in completing each of their allocated instruments on a visual analogue scale from 0 to 100.

Of the 62 eligible patients, 51 (82%) agreed to take part in the study and, of these, 42 (82%) completed at least three assessments of their allocated measures. Patients' ratings for the perceived validity of the instrument in measuring their quality of life were significantly higher for the SEIQoL-DW than for the SIP or SF-36. There were no differences in the perceived validity of the SIP or SF-36. Overall, the degree of emotional distress caused by the completion of any of the instruments was low, although values were only reported graphically and absolute numbers were not reported. The SEIQoL-DW had the lowest rating for emotional distress, followed by the SF-36 and then the SIP, with the difference between the SEIQoL-DW and the SIP being statistically significant.

This was a small study of patients with a relatively rare condition, and as such the findings may not be generalisable. Nonetheless, the study provides useful data for theory testing. In terms of the theories under test, these findings can be interpreted in two ways. First, they indicate that patients preferred the freedom to nominate areas of life important to them as a result of the open structure and content of the SEIQoL-DW, compared with responding to the pre-defined set items and response alternatives mandated by the standardised nature of the SF-36 and SIP. The authors also point out that the high distress ratings of the SIP may also be because the items focus on what patients are no longer able to do, rather than what they can do, and might have reminded patients about the progressive nature of their condition. Alternatively, as the authors point out, the differences in perceived validity and emotional distress ratings between the three instruments may be due to the fact that the SEIQoL-DW was interviewer administered and this interaction was preferred by patients and also served to reduce the emotional distress of instrument completion. These explanations are not mutually exclusive and both may account for the differences in the ratings between the three PROMs. However, this study did not collect any qualitative data to further explore how and whether these explanations work together or independently of each other.

Summary of quantitative findings on patient views of patient-reported outcome measure format and structure

The studies reviewed in this section suggest that 'not all PROMs are equal', and that they vary in how easy they are to understand and respond to and in the extent to which patients feel that PROM results reflect their health. We found some evidence to indicate that patients are prepared to tolerate a longer questionnaire that is more complex to complete if it provides a more comprehensive assessment of their health.³²⁸ However, PROMs that are too long and/or contain items that remind patients of how little they can do cause them distress.^{329,331} We also found some evidence to support the theory that patients view individualised measures as more valid and less distressing to complete.³²⁹ This may be because they prefer the freedom to nominate areas of their life that are important to them. It may also be because they value the interaction with an interviewer, which could facilitate the exploration of these areas and mean that patients feel that they are being listened to which, in turn, may reduce any emotional distress caused by PROM completion.

Qualitative evaluations of standardised and individualised patient-reported outcome measures

The studies reviewed so far have largely used quantitative methods to evaluate patients' perceptions of PROMs. We now review studies that have taken a qualitative approach to explore how meaning is constructed as patients frame their own health experiences in response to standardised and individualised PROM items. A basic assumption of PROMs is that the standardisation of items and content gives rise to a standardisation of meaning, thus enabling change in individual patients to be measured over time. The studies reviewed here delve into the 'black box' of how patients complete PROMs to test this theory.

Mallinson³³¹

At the time this paper was written, few existing PROMs had been developed with input from patients. Consequently, although PROMs had ample quantitative evidence to support their reliability and validity, few instruments had been subjected to qualitative evaluations to explore whether or not patients ascribed the same meanings and interpretations of items that were intended by PROM designers. 'Cognitive testing', that is, exploring how respondents interpret and understand items, is now advocated as an integral part of the development of new PROMs.²⁶⁸ However, many of the older standardised PROMs currently in use have not been subjected to this process, and therefore this study still has relevance today. Mallinson³³¹ acknowledges that differences between and within respondents in their interpretation of items may be 'ironed out' when such measures are used to compare groups within RCTs. Nevertheless, as Mallinson points out, the valid use of such measures in the care of individual patients 'rests on shared meanings or at least stable meanings within the individual across time and place'.³³¹

To explore these issues, the SF-36 was used as a case study. Fifty-six people aged ≥ 65 years who were newly referred to community physiotherapy and occupational therapy teams in the north-west of England were recruited to the study. Mallinson interviewed patients to administer the SF-36 as specified in the SF-36 user manual,²⁵² using standardised questioning. These interviews were recorded and analysed to explore respondents' spontaneous contributions to the interview process as evidence of 'troubles' in the sense-making process. The paper focuses on respondents' difficulties in answering two subscales of the SF-36 in particular, the 10-item physical functioning scale and the 5-item general health scale.

Analysis of the ways in which respondents approached answering these items and their spontaneous contributions revealed a number of technical difficulties with item wording and construction, but also problems that questioned the conceptual basis of the items and their implicit assumptions about the 'linearity and stability in peoples' [sic] conceptions of health'. The technical difficulties experienced by respondents included difficulties in responding to double questions (e.g. an item referring to 'bending, kneeling or stooping'), problems understanding unfamiliar words (e.g. knowing how far a mile is in practical terms in relation to familiar benchmarks such as the walk to the local supermarket) and struggling to make sense of vague questions (e.g. understanding bathing as meaning having a bath, rather than also incorporating having a wash or a shower). One could argue that these challenges could be addressed by careful rephrasing and rewording of items. However, the conceptual difficulties experienced by respondents posed more fundamental challenges to the premise of using standardised questions to capture respondents' views about their condition and its impact on their health. It is beyond the scope of our synthesis to review all the findings in detail here, and so we focus on two issues of particular relevance to the use of PROMs in the care of individual patients.

The first issue is that questionnaire items contain within them implicit assumptions about what is 'normal' that may not be shared by respondents. Consequently, items may not adequately capture the patient's level of health or changes over time. For example, Mallinson found that the normative level of physical activity implicit within the physical functioning scale was 'way beyond the capacity of most participants in this study'.³³¹ As such, the scale failed to capture the respondents' current levels of physical activity and the small improvements they may make in the future were unlikely to lift them any higher than the most severe response category, and, as such, the measure would fail to reflect these improvements. Mallinson also observed that respondents found it demoralising that they were unable to do anything implied as being

'normal'.³³¹ Furthermore, she noted that respondents felt that these items were irrelevant to them and displayed signs of irritation, boredom and withdrawal from the process of responding to the questionnaire. In these circumstances, it is less likely that respondents' answers will accurately reflect their experiences.

The second issue related to the process through which respondents answered the questions exploring their general health perceptions. One item asks respondents to rate their general health and choose from excellent, very good, good, fair and poor. To answer this question many respondents compared their own health with that of a reference group. However, respondents varied in the reference group used (other people their own age, themselves before they became ill, people with different conditions perceived to be worse than their own). Furthermore, choice of reference group had a substantial impact on their response, as one interviewee explained: 'There are people worse, so I'd say fair. But you see again, if somebody is dying of cancer I'm excellent but compared with you I'm very poor.'³³¹ If patients do not consistently use the same frame of reference over time, the questionnaire will either over- or underestimate changes over time.

This was a small but detailed study that opened up the 'black box' of the process through which patients responded to standardised PROMs and revealed the troubles therein. One could argue that these challenges can be addressed by ensuring that the PROM used is appropriate to the patient's level of health, for example by using computer-adaptive testing such as PROMIS and by providing patients with more guidance to ensure their frame of reference is consistent over time. However, perhaps most relevant to our synthesis is what this study reveals about the PROMs completion process overall. The PROM provided a structure into which patients had to fit their experiences and through which patients actively constructed their answers. This structure had the potential to distort the representation of patients' subjective experiences, rather like trying to fit a square peg into a round hole. Furthermore, the fundamental underlying assumption of standardised PROMs, that standardisation of question items and responses leads to standardisation of meaning, is challenged by the findings of this study. These observations hold whether the questionnaire is completed independently (e.g. when a patient completes a PROM prior to the consultation) or through an interview (e.g. when the PROM is completed during the consultation). Moreover, when a PROM is completed during an interview, as would be done if a PROM is completed during the consultation as a form of assessment, an additional layer of social interaction is brought into play. As Mallinson notes, in this context, standardised PROM completion is unlike the usual flow of conversation. The direction of questioning is one-way and the wording of the questions should not be altered, otherwise the validity of the PROM, as underpinned by psychometric testing, is threatened. Thus, as Mallinson notes, the standardised survey interview creates an 'interactional strangeness' where 'most of the mechanisms to check meaning are suppressed'.³³¹ In other words, the standardised approach required to uphold the validity of the PROM within a quantitative framework of psychometrics also threatens its subjective validity to patients.

Westerman et al.³³²

This article reported on a qualitative study that explored the process of cue elicitation between interviewers and patients with small-cell lung cancer while completing the SEIQoL-DW. In the process of cue elicitation, patients are read a 'script' that provides a definition of quality of life and are invited to nominate up to five areas that determine their quality of life by 'thinking about the areas of our life that would (or do) cause us most concerns when they are missing or going badly'.³³³ Data were obtained from interviews with 31 patients and included cues and notes recorded by the interviewer on the 'Cue Definitions Record Form', transcribed recordings of the SEIQoL-DW elicitation process and field notes made following these interviews.

In their findings, the authors focused on the different ways in which the interviewer was involved in coconstructing the respondents' answers to the SEIQoL cue elicitation process. The authors found that 12 patients were able to nominate five cues spontaneously, and that 13 had to be prompted with a list. Four patients were either too tired or too distressed to continue after eliciting one or two cues. Of the 26 patients who completed the SEIQoL, 17 nominated cues with little explanation. However, nine patients discussed their cues in a more elaborate narrative, which required considerable work on the part of the interviewer to translate the narrative into five cues. The authors observed that on some occasions the interviewer was actively involved in reorganising and combining respondents' cues. As such, the authors

noted that cue elicitation is not a one-way process whereby the patient nominates cues and the interviewer simply records these. Rather, the cue elicitation process involves the coconstruction of cues through the patient–interviewer interaction. As the authors noted, in producing such cues there is a tension between ‘the patients’ answers and the interviewers’ instructions’, during which the interviewer is trying to balance the ‘freedom’ of patients to tell their own story with the ‘control’ required by the instruction manual to reduce this narrative down to five cues that can then be rated. They remark that this requires a whole series of complex decisions on the part of the interviewer, which may be vulnerable to the introduction of unnoticed ‘bias’ because of differences in the ways these decisions are taken by different interviewers.

The authors also highlighted areas in which the process of reducing patients’ narratives down to five cues can result in the loss of meaning as it is recorded on the Cue Definitions Record Form. They noted that although both ‘staying positive for my family’ and ‘I’m happy with the support from my family’ could both be recorded as ‘family’, they have very different meanings. Furthermore, cues categorised in different areas may have similar meanings. They argued that it was questionable whether the meaning of such cues could be retained consistently over time and across interviewers. In terms of the theories under test, this study suggests that, although the process of cue elicitation in completing the SEIQoL-DW is intended to allow patients to nominate their own areas, in reality, the selection of cues is coproduced through the interaction between the patient and the interviewer. Variations in the way this selection process is carried out may lead to the definition of different cues within patients and across time. Furthermore, there is still the risk that meaning may be lost or distorted as patients’ narratives are reduced to five cues.

Farquar et al.³³⁴

This study explored all three stages of the SEIQoL-DW completion process in 13 patients with chronic obstructive pulmonary disease who were enrolled in a pragmatic RCT to explore the impact of a complex intervention for breathlessness in patients with advanced chronic obstructive pulmonary disease. Completion of the SEIQoL involves participation in a partially scripted semistructured interview. In the first step, patients are asked to nominate up to five cues that reflect areas important to their quality of life, in step 2 patients are asked to rate their current functioning in each of these areas and in step 3 patients are asked to rate the importance of each area nominated. The patient’s cues, their ratings for steps 2 and 3 and their fatigue or boredom with the procedure are recorded in an interview record form. Patients completed the SEIQoL-DW between three and five times each during the trial. Data comprised audio recordings of these interviews and completed interview record forms. In total, 48 complete recordings and calculations of the SEIQoL-DW scores were made. The authors identified a number of issues with completion of the SEIQoL-DW; we do not cover all the issues here but report those most relevant to our theories.

During step 1 (cue elicitation), the authors noted that some respondents were reluctant to nominate cues they felt unable to change, irrespective of how important they were to their quality of life. Others ‘forgot’ to include certain cues, despite discussing them during the interview. These findings suggest that the five nominated cues included in the SEIQoL may not always reflect issues that are important to the patient, although patients are provided with the opportunity to discuss and share them during the process of cue elicitation. The authors also observed that some patients did not spontaneously mention five areas and the interviewers felt a ‘need’ or ‘pressure’ to help patients identify cues.

During step 2, when patients are asked to rate their functioning in each area, the authors observed that participants held different interpretations of the phrase ‘how each of these areas are for you’, including ‘physical ability to carry out a cue’ and ‘satisfaction with ability to carry out a cue’. These different interpretations affected their ratings. For example, one patient had nominated ‘the process of going to bed’ as a cue; although the patient could do this, he found it burdensome, as it involved managing a lot of complex equipment. Thus, basing his rating on his physical ability to do the cue would have resulted in a high score, whereas basing his rating on his satisfaction with his ability to do the cue would have resulted in a low score. Similarly, during step 3, some respondents had difficulty understanding the instruction to ‘rate

the importance' of each of the five cues; some respondents rated this in terms of their current functioning on the cues and, as such, did not distinguish the difference between steps 2 and 3.

Again, this was another small but detailed qualitative that which opened up the 'black box' of PROMs completion. In terms of the theories under test, this study suggests that completion of the SEIQoL-DW may also result in some distortion of patients' experiences, as the five cues selected may not reflect the issues of most importance to the patient's quality of life. This is a function of the requirement to reduce the patient's narrative into five areas in order to rate them and produce a score. It also raises an important distinction between the *process* of completing the SEIQoL-DW, in which the patient may raise a number of issues, and the resulting cues and scores of the SEIQoL-DW, which quantify the patient's HRQoL so that it can be used to measure outcomes over time. If the instrument was administered by a clinician, the clinician would have access to both the patient's narrative and the score. However, if the SEIQoL was administered by an interviewer and the score was fed back to the clinician, the clinician would have access only to the cues and scores, which may not fully reflect the patient's areas of concern. This study also indicates that the standardisation of phrases used in the process of rating the SEIQoL, like those in Mallinson's study of the SF-36,³³¹ did not lead to a standardisation of meaning. Patients interpreted seemingly simple phrases in different ways, which had the potential to affect their ratings.

Summary of qualitative findings on patient-reported outcome measure format and structure

Mallinson's qualitative evaluation of the SF-36 suggests that the structured nature of standardised PROMs may result in a distorted representation of patients' perceptions of their health.³³¹ This study also challenged the fundamental assumption that the standardisation of items leads to the standardisation of meaning; it cannot be assumed that PROMs items hold a shared meaning both within patients over time or between patients and clinicians. This finding has important implications for the use of routine collection of PROMs data to inform the care of individual patients. It suggests that in order for individual PROMs data to put 'patients and clinicians on the same page'³³⁵ during a clinical consultation, clinicians and patients will need to engage in a process of 'sense checking' PROMs data, to ensure that both parties have a shared understanding of its meaning in order to inform discussions and decision-making about care and treatment. This study also highlighted that, where PROMs are used to assess patients during a consultation, the standardised approach to questioning that is required to uphold the validity of a standardised PROM may suppress this sense-making process and, as such, lead to the production of patient responses that may not reflect their perceptions. These findings pose important questions about how patients and clinicians manage this process when individual PROMs data are collected and fed back to clinicians to inform patients' care.

We also found evidence to suggest that the resulting scores produced by the completion of individualised PROMs such as the SEIQoL-DW may also present a distorted representation of patient's perceptions of their quality of life^{332,334} as a result of the requirement to reduce patients' narratives down to five cues and then rate their current severity and importance: in other words, to *quantify* them. Furthermore, standardised phrases in these instruments were also interpreted differently by respondents and further demonstrate the argument that the standardisation of questions does not necessarily lead to the standardisation of meaning. However, unlike standardised PROMs, the completion of the SEIQoL-DW is a semistructured rather than a structured interview. Under these circumstances, there is more scope for the meaning of items and phrases to be negotiated between respondents and the interviewer, although this process may bring its own challenges, as the interviewer may introduce 'unnoticed bias' into the ways in which this meaning is negotiated. Furthermore, this process enables patients to tell their own narrative, even if this is then reduced to five cues. This raises an important question about the function of individualised measures and their role as a 'conversation opener' about the patient's quality of life versus their role as an 'outcome measure' to quantify this quality of life and measure changes over time. The findings of these studies suggest that the individualised measures may have more value as a 'conversation opener' than an outcome measure, especially if they are administered by a clinician, as they invite patients to identify issues that most affect their quality of life in their own words. This is a theory that can be tested empirically.

Reviewing context configurations: how do they shape the intended impact of patient-reported outcome measures completion?

With one exception,³²⁵ the above studies have examined how patients approach PROMs completion outside the context of clinical practice. We now review a series of studies that examine patients' and clinicians' views of the extent to which PROMs are easy to complete, capture patients' views and enable the sharing of information in real clinical settings. We first explore their use in the care of people with mental health problems. In primary care, the use of standardised depression questionnaires was incentivised under the QOF between 2003 and 2013. In secondary care, the collection of PROMs in both adult and child mental health services has been advocated as a means of both supporting the care of individual patients and monitoring the quality of services for this population of patients. For example, in England, the Increasing Access to Psychological Therapies in both adult and child and adolescent mental health services (CAMHS) programme advocates the routine collection of PROMs, to support the care of individual patients and at the same time to monitor service quality; however, it has not been incentivised with monetary rewards and sanctions. In primary care, patients are more likely to have an existing relationship with their GP. In secondary care, some patients are newly referred and others may be returning patients, but it is less likely that they will see the same clinician. Thus, with each new episode of care, clinicians and patients must build a relationship, which is also seen as instrumental to the success of therapy.

Patients' and clinicians' experiences of using patient-reported outcome measures to detect and manage depression in primary care

We begin by reviewing studies that have explored the use of PROMs to support the care of people with depression in primary care.

Dowrick *et al.*⁸¹

This study examined views on the use of standardised questionnaires to assess depression among UK patients and GPs while their use was still incentivised under the QOF. Interviews were conducted with 34 GPs and 24 patients. These participants were recruited from 38 general practices around Southampton, Liverpool and Norwich that took part in a quantitative study on depression severity questionnaires. In the selection of GPs, variation was sought in terms of gender, years of experience, full-time/part-time practice, trainer/non-trainer status, geographical location and size of practice. For patients, selection considered sex, age, ethnicity and sociodemographic background.

Patients were largely positive about the questionnaires, seeing them as an efficient approach and an indication that clinicians took their problems seriously. As one patient explained, 'It can be perceived as you're being taken more seriously'.⁸¹

The findings suggest one way in which PROMs may encourage accurate description of symptoms by patients: making it easier for them to record their concerns by providing a less threatening format. One patient explained that the questionnaire helped to indicate their concerns because with some issues:

[Y]ou can't say how you feel and with answering the questions, it, like, asked you how you felt and things like that and now I think that helped, just like, ticking the boxes, just so that he knew . . . how I felt as well.

PT21, 20–25⁸¹

Other patients commented that the questionnaire had a less personal nature, which made it easier to answer the questions and share information with their doctor, as one patient explained:

[I]f it had been that the doctor had asked you the questions . . . then it might be a bit different. I think the impersonal nature of the questionnaire's probably helpful.

PT11, 107–10⁸¹

The authors reported that some patients saw the PROM as an opportunity for reflection, which in turn helped them to communicate their concerns to clinicians more effectively. Some patients reported that the questionnaire had helped to increase their self-understanding, so that they could organise their thoughts and express themselves better to their doctor, as one patient reported:

I think that [completing the questionnaire] helped me in my head as well . . . Well I started to think, you know about why I was getting depressed and that.

PT19, 45–48⁸¹

However, the study pointed to different views among patients on the PROMs format. One patient felt that the scoring system was too reductionist and did not allow her to accurately express her status:

I suppose, yeah, it sort of quantifies that you are . . . you do have problems . . . but I, I still feel that it was, like . . . you're trying to, like, tie a number to a thing which isn't necessary, isn't necessarily like a yes or a no . . . it's a, it's very difficult to . . . put a description to it, I think.

PT03, 98–101⁸¹

This suggests that although a closed format may help some patients to express their feelings, other patients may see this format as a barrier, such that the PROMs questionnaire may not allow what they see as an accurate picture of their health status. The authors also suggested that some patients may not complete the PROMs accurately owing to concerns about the stigma attached to depression. As one patient explained:

You're more likely to lie, well I found I'm more likely to lie . . . Because I still find a lot of stigma attached to depression.

PT16, 100–5⁸¹

In addition, some patients reported manipulating or 'gaming' their PROMs feedback to influence the care they received:

I said no to anti-depressants as I wasn't depressed. And then she gave me a questionnaire, I remember quite clearly . . . And I remember reading . . . and I thought, I'm not putting down how I really feel, and so I didn't.

PT24, 131–42⁸¹

These themes seem to have been reported by only a small number of patients within the study sample and it is unclear how widely these views are shared.

General practitioners were more cautious and had concerns about the validity and usefulness of the questionnaires. Furthermore, GPs also reported some manipulation of the data by avoiding coding people with depression to avoid having to administer the depression questionnaires.

This study provides useful evidence about conditions that can affect patient reactions at this stage of completing the PROM, and, through this, whether or not PROMs record accurate information. Based on the findings, it appears that standardised depression questionnaires can make it easier for patients to disclose their concerns, when patients feel more comfortable admitting to sensitive issues through ticking a box rather than articulating concerns verbally. However, some patients may feel that a standardised format does not allow accurate description of their concerns. Second, patients may still not complete the forms openly if they are embarrassed about disclosing sensitive problems, suggesting that the relative ease of ticking a box does not always overcome stigma. In addition, patients may intentionally provide inaccurate information if they know or are concerned that their answers will have unwanted implications for their care.

Leydon et al.³³⁶

This study³³⁶ was a further analysis of the interviews conducted with the 34 GPs as reported by Dowrick et al.⁸¹ Here, the analysis focused on GPs' views of the impact of standardised depression questionnaires on the doctor–patient relationship and the flow of the consultation. As such, the paper is particularly relevant to the theory that the use of standardised PROMs may interfere with the doctor–patient relationship.

The findings indicated that doctors placed a lot of emphasis on developing a relationship with patients in order to detect and manage depression. One GP explained that the 'one to one personal relationship' was especially important when they were 'supporting someone with depression', while another observed that 'a lot of picking up depression is about rapport and about patients feeling comfortable and establishing a relationship'.³³⁶ This relationship building was achieved through talking to patients and letting patients express themselves in their own words. GPs reported feeling that standardised depression questionnaires threatened this relationship-building process by covering issues that had already been discussed, but in a mechanistic way, and trivialising the strength of patients' emotions, as the quotations below illustrate:

Getting them to do the tick box thing at the end of all that, when they've already bared their soul, can sometimes be a bit . . . intrusive I find.

GP³³⁶

If you've had a very loaded consultation, very cathartic . . . the HAD [Hospital Anxiety and Depression] scale can appear to trivialise the depth of emotions that are being expressed.

GP³³⁶

General practitioners also found it difficult to integrate completion of the questionnaire during the consultation, as one GP explained: 'the greatest challenge is how to incorporate it tactfully into the consultation'.³³⁶ GPs reported that they felt uncomfortable using a standardised questionnaire during the consultation when a patient first came to see them about depression, as the two quotations below show:

. . . when a patient comes with a first presentation of depression . . . I don't feel comfortable, straightaway with . . . issuing a box-ticking exercise.

*Where do you plonk those great big . . . bombshells in the middle of a normal consultation with somebody.*³³⁶

These quotations imply that the administration of depression questionnaires was not part of the 'normal' activities one might find during a consultation. Instead, the use of words such as 'box ticking' and 'bombshell' suggest that the questionnaires were viewed as bureaucratic and destructive to the normal flow of conversation. GPs had developed a number of strategies to manage this conflict, including giving patients the questionnaires to complete at home, as one GP explained:

*[W]hat I would tend to do is give them a form and say, look, take it away at your leisure.*³³⁶

This study suggests that GPs felt that developing rapport and a trusting relationship with the patient was essential to enable patients to share their feelings and emotions in order for the GP to detect and then support people with depression. In terms of the theories under test, the use of standardised depression questionnaires threatened this process because these questionnaires appeared to revisit issues that had already been discussed and to trivialise patients' emotions. They were not seen as vehicles for enabling patients to share their emotions and feelings; rather, they were seen as bureaucratic and destructive. Furthermore, it was difficult to incorporate the questionnaires into the consultation without disrupting the flow of conversation and, thus, the relationship-building process. At the same time, the use of these measures were incentivised under the QOF, so rather than avoiding their use altogether, GPs had developed 'workarounds' which included avoiding coding people as having depression (as reported by Dowrick et al.⁸¹ on the basis of the same interview data), so that the use of the questionnaires was not necessary, and giving patients the questionnaires to complete at home.

Mitchell *et al.*⁸⁰

This study also explored primary care practitioners' views of the use of standardised depression questionnaires in the diagnosis and management of depression in primary care in England. The authors took a purposive sample of four GP practices and conducted one focus group in each practice. Focus group participants included GPs, practice nurses, community nurses, primary care mental health workers and practice managers. Participants were asked to describe how the introduction of the QOF and NICE guidelines, with their requirements for GPs to use standardised depression questionnaires, had influenced how depression was diagnosed and managed.

The authors reported that GPs found the use of the depression questionnaires 'counterintuitive, intrusive and unnecessary'.⁸⁰ One GP explained that:

*they . . . break down in tears and tell you how depressed they're feeling . . . and you ask them questions and then 'oh now I've got this questionnaire to fill out'. I just think it's so inappropriate sometimes.*⁸⁰

This same GP also felt that the standardised questionnaire was often unnecessary because:

*it's not going to change your management because it might be someone you know quite well but you have to . . . get the QOF points.*⁸⁰

In other words, clinicians did not see the questionnaire as essential to the management of the patient, as they were already familiar with their history, but they felt obliged to use the questionnaire because the practice would lose QOF points, and thus income, if they did not.

Consequently, the authors observed that GPs had adapted how they administered a particular questionnaire, the PHQ-9, to fit with their consultation style. These included some methods which did not compromise the validity of the questionnaire, including letting the patient self-complete the measure in the waiting room and administering the measure via telephone. However, other 'workarounds' may have compromised the validity of the PHQ-9. These included GPs reading out the questions to the patient and recording the answers themselves, missing out the last item on the PHQ-9 because it was not necessary for the QOF points, recalling the questions from memory during the consultation and working out the score afterwards, and adapting the wording of items via a translator for patients whose first language was not English.

Pettersson *et al.*³³⁷

This study aimed to explore the views of Swedish GPs regarding the use of standardised depression scores to detect and manage patients with suspected depression in primary care. In Sweden, the use of standardised depression questionnaires in primary care has been recommended by national guidelines, but no monetary rewards or sanctions or routine data collection have been mandated. GPs were recruited from an urban area (Gothenburg) and a medium-sized town (Skövde); GPs in Gothenburg were also involved in a trial to examine whether or not the routine collection of a standardised depression rating scale improved recovery in patients with mental health problems. Of the 30 GPs invited to participate, 27 took part in one of five focus groups. GPs were asked to discuss when they did and did not use the instruments, how the use of the instruments affected the consultation and the extent to which they thought that the questionnaires provided an accurate representation of the severity of a patient's depression. All focus groups were audio recorded and field notes were taken.

The authors found that all GPs emphasised the importance of detecting and managing patients with depression based on their own experience; instruments seldom added any value. As one GP explained:

With questionnaires, it feels like going back in development and starting at a more basic level where you give every detail the same importance although you in reality can drop most information rather quickly and concentrate on a few things in order to get a clue.

M2 FG1³³⁷

They felt that the use of standardised depression scores may lead to a more bureaucratic work style and were concerned that this may limit a GP's ability to think for him- or herself.

General practitioners suggested that the PROMs distracted them from noticing a patient's non-verbal communication, such as body language, which was as important to GPs in understanding the patient's feelings. As such, PROMs completion limited communication between the doctor and the patient. As one GP explained:

[...] misses the opportunity to pick up that little extra that was of special importance to the patient, to hear the nuances in what is said and not said.

Female 7 (F7), FG2³³⁷

Most participants perceived that the consultation was negatively affected when an instrument was used because it constrained the narrative of the patient and interfered with the GP's ability to listen to the patient. GPs perceived that PROMs limited the patient's opportunity tell their own unique life story. Furthermore, GPs viewed that this affected the rest of the consultation:

Because then it is said and done and we should of course follow up and have a normal consultation talk but in some way both parties are locked by this protocol.

M7, FG3³³⁷

General practitioners also felt that the use of standardised questionnaires restricted their ability to be active and engaged listeners. They observed that patients expected to receive their full attention and to engage in a dialogue and a discussion. GPs perceived that using a standardised depression questionnaire led them to feel 'alienated' from patients' problems. They also felt uncomfortable with trying to reduce patients' emotions to a score.

Many GPs were uncertain whether standardised questionnaires provided a valid and reliable representation of the patient's problems. This was for a number of reasons. Many expressed the view that questionnaires could not capture the complexity of such a condition as depression. They also reported that a number of patients had misunderstood the questions and had then gone back and changed their ratings once the GP had explained what the questions meant. We can hypothesise that such help might have compromised the validity of the questionnaire, but not giving help would have rendered the results of the questionnaire meaningless. GPs also questioned the validity of patients' self-ratings, which they felt would vary depending on the patient's life situation. One GP observed:

I am convinced that we have patients that always score very high on MADRS-S [Montgomery-Åsberg Depression Rating Scale] and yet live their normal lives. But they only go to the doctor the day they can't handle their work.

F7, FG2³³⁷

Although GPs felt that most patients completed the measures truthfully, some raised concerns that some patients may game or artificially inflate their scores to obtain sick leave:

... but I think that we may also be manipulated. Symptom scales are available on the Internet and you have threads on Flashback on how to proceed in order to get a sick leave.

M5, FG2³³⁷

However, all GPs agreed that questionnaires were useful in some situations, for example with patients who found it difficult to express themselves. For example, one GP related a case of a patient who only gave 'weak signals' that she might be depressed, but, on completing the questionnaire, it became clear to both the patient and the GP that she was depressed. Some GPs used questionnaires after the consultation to check that they had not missed any important aspects. The ratings of various items were often used as

starting points for discussion, thereby improving communication. They were also found to be useful for patients who could not accept that they had a psychiatric diagnosis:

Men in their forties can be very hard to convince. They prefer to have an ulcer diagnosis.

F5, FG1³³⁷

This was a small qualitative study of a self-selected sample of GPs in two areas of Sweden. It is possible that those with strong feelings (either positive or negative) were more likely to have taken part in the study. However, it provides another layer of evidence to suggest that GPs felt the use of standardised depression questionnaires hampered their relationship with the patient and constrained the patient's ability to tell their story in their own voice. They also questioned the ability of depression questionnaires to adequately and accurately capture the patient's problems.

Summary of findings from managing depression in primary care

These four papers provide a useful synthesis of the complex ways in which different drivers and incentives may support or constrain the use of PROMs data by different stakeholders for different purposes. It must be noted here that the majority of the evidence reviewed in this section reflects the views of GPs, who, some might argue, may have felt threatened by the imposition of these measures. It is possible that those who agreed to be interviewed in these studies were simply those who held the strongest opposing views to the use of PROMs in primary care.

The patients in Dowrick *et al.*'s⁸¹ study were, not surprisingly, unaware that the use of the measures were incentivised within the QOF. For some patients, completing a standardised questionnaire was a less intrusive way of sharing their feelings with their doctors and also enabled a process of self-reflection which helped them to understand why and how they were feeling depression. Other patients felt that the standardised nature of the PROM did not fully reflect the complexity of their experiences. However, some patients suspected that PROMs may be used to limit their access to care and, under these circumstances, they admitted misrepresenting their experiences in order to avoid this outcome.

In contrast, GPs felt that it was essential to develop a rapport and a trusting relationship with the patient, so that the patient felt able to share their feelings and the GP could then identify and support those patients who had depression. GPs perceived that using standardised depression questionnaires constrained this process by trivialising patients' emotions, and they found it difficult to incorporate the use of standardised measures into the consultation. Although not explicitly mentioned by GPs themselves, their comments referring to the use of standardised depression questionnaires being a 'tick box' and 'bureaucratic' exercise echo Mallinson's³³¹ observations about the 'interactional strangeness' of standardised surveys, where opportunities for sense making and relationship building are constrained.

General practitioners in England found themselves in a difficult position. They did not perceive that standardised depression questionnaires were helpful in supporting their care of individual patients with depression. However, because the use of such measures was incentivised under the QOF, they stood to be penalised if they did not use them. Under these circumstances, GPs developed a number of strategies to meet these two conflicting drivers, by changing the ways in which they administered the standardised depression questionnaire. Some of these methods, which attempted to mitigate the difficulties they experienced in incorporating standardised questionnaires into the consultation, threatened the psychometric validity of the questionnaire, thus defeating its value in the care of individual patients. Furthermore, GPs were also reluctant to code patients as having depression, to avoid being penalised for not using a standardised depression questionnaire, which challenged the measure's value as an indicator of the quality of care for people with depression. Here, we see an example that attempting to use the same PROM to support the care of individual patients and as an indicator of the quality of patient care can result in PROMs being used in a way which undermines their use for one or both purposes.

Patients' and clinicians' experiences of using patient-reported outcome measures in secondary care mental health services

We now consider how PROMs have been used in secondary care to support the care of adults and children with mental health problems. Here, the collection of PROMs has been advocated as a means of both supporting the care of individual patients and monitoring service quality. However, there are no monetary rewards or sanctions attached to the collection of these data. The majority of studies here focus on the use of standardised PROMs, but one study³³⁸ explored the use of an individualised PROM in drug and alcohol services.

Hall *et al.*³³⁹

The Children and Young People's IAPT programme is an initiative delivered by NHS England that aims to improve both access to and the quality of CAMHS. A key component of this is the use of session by session outcome monitoring to inform the individual care of service users. To inform the choice of measures for this purpose, this study reports on a study that piloted a session-by-session outcome monitoring tool in three UK CAMHS clinics. The monitoring tool was SxS, an electronic 8-item questionnaire based on the Strengths and Difficulties Questionnaire. Patients and/or their parents or carers completed the questionnaire on a tablet computer in the waiting room before each appointment. The tool automatically created a report graphing the patient's progress, which was provided to the clinician, the idea being that this would be discussed during the appointment. In this context, the SxS was expected to serve two purposes: the care of individual patients was informed through feeding back individual patient data to clinicians but also data could also be aggregated to monitor the quality of the service. Semistructured interviews were conducted with clinicians ($n = 10$), administrative staff ($n = 8$) and families ($n = 15$) to elicit their views on the monitoring tool.

Clinicians made general comments on their attitudes towards outcome measurement, which suggested that they did not find outcome measures useful for service development because they did not receive any feedback on this aspect of data collection. As one clinician explained:

*they are important for the service . . . but . . . I don't ever see any data coming back from them . . . so I don't think they are very useful . . . it would be more useful if we were getting feedback from them.*³³⁹

In this study, patients completed the questionnaire prior to their session with the clinician. Clinicians felt that this was the most appropriate time to complete the measure. Completing the measure during the session could 'sometimes just get in the way of the therapeutic process a bit'.³³⁹ Completing the measure after the session would not enable data to be used with patients during the session and could also resulted in flawed data for service monitoring purposes because patients may feel 'like they are giving us feedback' instead of reporting on the symptoms and thus could 'feel under pressure to say something . . . more positive than they really think'.³³⁹

Patients and clinicians held different views of the benefits of PROM collection. The authors observed that families felt that 'completing the questions was a helpful process in its own right',³³⁹ because it provided an opportunity to reflect and assess their condition, which made them more prepared for discussion with clinicians. The authors further noted that:

*Parents and young people reported that it reminded them to ask specific questions to their HCP [health-care provider] and that it gave them an opportunity to reflect on how they had been since their last session. As a result of this, families reported feeling more prepared for their session.*³³⁹

Some patients also commented that completing the measure showed them how far they had progressed over time and that this provided the motivation to keep working on themselves in therapy. As one young person explained, 'It showed you . . . how far you've been going and . . . giving you motivation to try and get it up another step'.³³⁹ However, this was possible only if clinicians discussed the findings of the measure with them; otherwise, they felt that there was no point to completing it, as one young person observed: 'I think there's not really much point in sitting and doing a questionnaire that isn't being used to

help you.³³⁹ However, some clinicians felt that referring to the measure results in the session could create awkward silences, during which the young people they were working with became distracted, as one clinician described: 'You've got to read it and then they sit there in silence and then you've got somebody who has got tics or has ADHD [attention deficit hyperactivity disorder] and they start kicking at blocks, the door, shouting . . .'³³⁹

This study provides some evidence in support of the theory that completing PROMs can help support patients to reflect on their problems and to raise their concerns with clinicians. It suggests that PROMs enable reflection that makes patients better prepared to raise issues and questions in consultations. It also suggests that patients value feedback on their progress and this can motivate their continued engagement with therapy. However, this is possible only if patients are also provided with a copy of the scores or if clinicians discuss them with the patient during the session. Clinicians reported that referring to the scores during the consultation sometimes interfered with young people's engagement.

Wolpert *et al.*³⁴⁰

This paper aimed to explore the views of clinicians and service users of CAMHS and diabetes services regarding the use of routinely collected PROMs data in England. Semistructured interviews were conducted with 10 participants from CAMHS (six young people and four clinicians) and 14 participants (four young people, seven mothers and three clinicians) from a children's diabetes service. Service users were recruited through advertisements in voluntary sector organisations and, as such, may not represent the range and diversity of service users of child mental health or diabetes services. This group had not experienced the routine collection of PROMs in their service and therefore interviews explored their views on two measures (a symptom checklist for CAMHS and the PeDsQL for diabetes services). During the interviews, participants were asked what information clinicians should seek in order to understand patients' priorities for care and their views on implementing the routine collection of PROMs with the services they used.

The young people using the diabetes service felt that explaining their experiences to clinicians was difficult and held mixed views about whether or not PROMs would support them in this process. One young person felt that completing a PROM may 'show what I feel like . . . so clinicians can help me', while another could not see how it would help her discuss her concerns: 'I put it down because it . . . matched how I felt but I'm not sure how to approach it now I have written it.'³⁴⁰ Participants also commented on the format of the PROMs; clinicians reported a 'real tension' between the fluid and unique nature of patients' views and the fixed and standardised format of PROMs. Service users also questioned if the fixed format PROMs could adequately capture the dynamic and continually changing nature of their experiences. For example, one young person with diabetes explained that she felt it would be difficult to complete a PROM before her consultation with the clinician because 'I might start doubting whether my answers were right . . . if I do it quickly, on the paper, I've done it and I can't go back and change it'.³⁴⁰ This quotation suggests that the process of completing the PROM gives a fixed and permanent status to their experiences at one point in time; however, when service users' experiences change rapidly, the PROM may not represent how they feel at the time of the consultation if it was completed days before.

The authors also noted differences between service user and clinician views on the timing of PROMs collection. The authors note that clinicians thought PROMs were most beneficial if collected immediately in order to target interventions to patients' needs and produce 'better . . . quicker outcomes'.³⁴⁰ In contrast, young people felt that they needed to 'build rapport' with clinicians 'instead of going straight to the nitty gritty'.³⁴⁰ Thus, before young people felt able to share some of the more sensitive information contained in the PROMs data they needed to develop a relationship with the clinician. As one young person explained, for some of the PROM questions, 'You'd tell your best friend or someone you were really, really close to but you wouldn't tell this random person you'd just met'.³⁴⁰

Participants also discussed concerns about how PROMs data would be used. A particular issue for the CAMHS service users was the implications of PROM completion for the service itself and their access to it. For example, one service user expressed concern that if she indicated on a PROM that she was worse than

last time then this would be 'rude', reflect badly on the service and imply that it has not helped her. Another was worried that the PROM would be used to restrict her access to the service: 'If I tick "much worse" then I don't know if [that means] "I don't need this service", I don't know, I'm very confused.'³⁴⁰ Parents of children with diabetes were concerned that focusing consultations on emotional issues might mean that less time was available for discussing diabetes management or the 'objective hardcore' and could bring unwanted professional scrutiny of the care they provided to their children.

This was a small study in selective group of clinicians and carers and parents of children who used CAMHS. Nonetheless, it adds a further layer of evidence to our understanding of how patients approach PROMs completion. In terms of the theories under test, this study suggests that the fixed nature of PROMs may not fully capture patients' experiences when those experiences change rapidly from one day to the next. This study indicates that services users have concerns about the use of PROMs when they are used to judge the quality of care they receive or restrict access to the services they receive. The study also suggests that the extent to which a PROM may enable or constrain the sharing of sensitive information with clinicians also depends on the patient's relationship with and trust in the clinician.

Stasiak *et al.*³⁴¹

This is another study that explored the views held by child and adolescent service users and their parents regarding outcome measures and their routine collection in CAMHS in New Zealand. The authors conducted nine focus groups that included 34 child and adolescent service users and 21 family members. In the focus groups, participants were invited to try out completing some commonly used PROMs and were asked to comment on their ease of use, acceptability and relevance. Family members were also asked about the possible difficulties of and concerns about outcome measure collection. The analysis conducted by the authors was largely descriptive, with little attempt to explain their findings. Here we reframe some of their findings in relation to the theories under test.

Participants commented on the process of completing outcome measures and the acceptability of completing the measures. Two key tensions emerged with particular relevance to our synthesis. First, there was a tension between wanting the measure to be brief and at the same time comprehensive enough to capture the complexity and changing nature of patients' symptoms. The authors report that young people felt it burdensome to complete long measures on several occasions and, as one young person explained, 'if it's too long, you're just going to make the answers up'.³⁴¹ However, at the same time, participants questioned whether or not measures fully captured patients' specific experiences or the dynamic and changing nature of patients' symptoms. As one young person explained, 'I can be really unwell and . . . get better and then I can get better and then I can get really unwell . . . so what's the point of having these forms?'.³⁴¹

The second tension was whether completing the measures supported or constrained young people's ability or willingness to express sensitive issues to clinicians. Key contextual determinants of this were the timing of completion and the degree of rapport and trust between the young person and the clinician. For example, the authors noted that young people felt completing a written measure to be 'easier and less embarrassing than talking to a clinician about difficult issues'. However, young people also reported that the questionnaires should be administered by someone with whom they had developed rapport. One young person explained, 'I think you'd have to have a trust thing built up first to actually share something with that person'.³⁴¹ The authors also argued that 'children and young people may not respond accurately to questions they regard as personal or sensitive unless there is good rapport with clinicians'.³⁴¹

Participants also discussed the positive and negative ways aspects of completing a measure and the uses to which data could be put. On the positive side, some parents felt that completing an outcome measure signalled that someone was taking their experiences seriously. As one parent explained, completing an outcome measure 'felt good because you finally had someone listening to you and trying to help you through what you were going through'.³⁴¹ The authors noted that young people felt that having the means to track their own progress was useful and 'potentially empowering'.³⁴¹ Young people also felt that

completing an outcome measure at the start of treatment provided a useful baseline 'so that everyone knows where you are' and that future progress could be measured against it.

However, young people and family members also expressed some concerns about the value of outcome measurement and the use to which data were put. Parents wanted reassurance that data were going to be used to inform the care of their child: 'is it going to change how you treat her, is it going to make things better, are you actually going to hear what we're saying?'.³⁴¹ They were also concerned that politicians might use data to restrict access to services. One parent thought that politicians might look at data collected and think 'we've "just got this nice little test that we've just proved that they've made X amount of progress" so that's it, end of service'.³⁴¹

This was a small study in a self-selected group of services users and families who used CAMHS. However, the findings echo many of those later found by Wolpert *et al.*³⁴⁰ In terms of the theory under test, this study suggests that the structure and format of PROMs creates a tension between the requirement for measures to be brief to maintain respondent engagement and the need for the measures to be detailed and flexible enough to capture the complexity and dynamic nature of service users' experiences. However, the study also suggests that it is not just the format of a PROM that supports or constrains patients' honest or accurate sharing of their experiences. Rather, it suggests that sharing experiences through a PROM are constrained when patients have not developed a trusting relationship with the clinician caring for them or when they are concerned that PROMs data may be used to restrict their access to services.

Cheyne and Kinn³³⁸

This study explored counsellors' views and experiences of using an individualised PROM, the SEIQoL, during the process of counselling service users with drug and alcohol problems. The SEIQoL was being evaluated as part of a controlled trial, where intervention group counsellors ($n = 3$) were trained to use the SEIQoL and were then expected to use it during their sessions with clients. During the 6-month trial, 20 clients received counselling using the SEIQoL. This study reports on findings from three focus groups with the counsellors who used the SEIQoL during counselling, carried out at 12 weeks, at 18 weeks and at the end of the study period. Data from these groups were used to inform the development of a questionnaire that was completed by all the counsellors. One of the research team also reviewed patients' case records to extract data relating to the use of the SEIQoL, services users' and counsellors' views of the SEIQoL and service users' responses to counselling. However, in reporting their findings, the authors do not always make it clear whether the data reported came from case notes, focus groups or questionnaires.

The authors report that the counsellors found the SEIQoL easy to use and that 17 out of the 20 clients fully understood the aims and use of the SEIQoL (three were uncertain). The case notes revealed that one service user experienced some boredom with SEIQoL, whereas the other 19 'were directly engaged by it'.³³⁸ The counsellors felt that the tool was flexible and fitted 'well in an assessment'.³³⁸ Counsellors' notes also captured services users' comments on the SEIQoL, which reveal that they found it useful to 'get a better picture of my life right now'.³³⁸ Counsellors felt that the use of the SEIQoL enabled service users to engage more willingly in self-reflection, which led to service users identifying areas of their life they wanted to tackle first. Both counsellors' own reflections and their impressions of reactions from service users illustrate this:

Helped the service user look inward and stop and reflect on whether there was another way of looking at their situation.

Counsellor³³⁸

I liked the subtle way it allowed the service user to get another perspective through their own thinking.

Counsellor³³⁸

It changed the way I saw drink screwing up my life.

*Service user*³³⁸

The disks helped me to weigh things up and choose what I wanted to deal with first.

*Service user*³³⁸

Counsellors perceived that because the SEIQoL supported the service user to engage in a process of self-reflection, this in turn facilitated the counselling process, through enabling the service user to 'talk more openly' and accelerating 'the learning and insight the service user had about their drink problem'.³³⁸ As one counsellor explained, the SEIQoL 'helped the service user to look at whether they were ready to change [and] . . . decide what was important to them and . . . whether they were confident to get on and do it'.³³⁸ Counsellors also felt that use of the SEIQoL prompted them to listen and reflect more on what the service user was saying. The counsellors also observed that use of the SEIQoL therefore helped them to 'form a better relationship and trust with the service users' and enabled them to 'get alongside the service user more quickly'.³³⁸ The counsellors attributed this to the fact the SEIQoL was 'person centred' and 'person led' and that it put the 'control back with the service user'.³³⁸

This was a very small study of the perceptions of only three counsellors in one service who were likely to be enthusiasts about the SEIQoL. The data are all gathered from counsellors and it is questionable how far their views reflect the feelings and experiences of their clients. The findings are overwhelmingly positive and read as a 'good news' story about the instrument. Nonetheless, the study provides useful insight into the mechanisms through which use of an individualised measure might enable service users to share their experiences with clinicians. In terms of the theories under test, this study provides some support for the theory that use of the SEIQoL facilitates the process of self-reflection and enables service users to prioritise their problems. In turn, this leads service users to engage more effectively with therapy and supports the relationship-building process with the counsellor.

Summary of findings from secondary care mental health services

These studies echo a number of the findings from the previous section examining the use of standardised depression questionnaires in primary care. They indicate that patients in secondary care also expressed concern that PROMs data would be used to restrict their access to services and that these worries increased the likelihood that they may 'game' their answers in order to avoid an unwanted outcome. They also suggest some ambivalence on the part of clinicians for the measures being used for multiple purposes; for example, clinicians did not see the measures as useful for monitoring service quality as they themselves received no feedback of these data.

The studies also suggest that PROMs completion may support patients to engage in self-reflection about their own condition, which can make them feel better prepared to see clinicians. There was also some evidence to suggest that PROMs completion may enable patients to track their own progress, which can motivate them to engage with the therapeutic process. However, this is only possible if patients also received feedback on their progress. The studies highlighted that longer, more comprehensive measures may better capture patients' experiences but may also lead to lack of engagement with the measure, making it more likely that respondents would make the answers up. This may be a particular issue for younger people. They also raised questions about the extent to which standardised PROMs can fully capture the complex and dynamic nature of patients' experiences. The studies also suggested that patients preferred to share information contained in PROMs when they had developed trust in the clinician caring for them. The study in which an individualised PROM was used suggested that its use supported this relationship-building process.

Clinicians' and patients' views of patient-reported outcome measures in palliative care

We now review studies exploring patients' and clinicians' experiences of routine PROMs collection in palliative care. The goal of palliative care is to maximise quality of life for those patients with incurable disease by controlling symptoms and addressing patients' psychological and spiritual concerns, suggesting that a systematic and comprehensive understanding of patients' quality of life is vital.³²³ For this reason,

the routine collection of PROMs in palliative care has been advocated and guidance has been produced on the selection and implementation of measures.³²⁴ However, the routine collection of PROMs within palliative care has not been part of formal government policy or guidance. It therefore provides a useful contrast to the use of PROMs in the care of people with mental health problems, which have been incentivised and become the subject of government-backed programmes.

Hagelin *et al.*³⁴²

This study investigated palliative care nurses' experiences of using a PROM within an inpatient palliative care service. Nurses were expected to give patients the European Organization for Research and Treatment of Cancer Quality of Life (EORTC QLQ-C30) questionnaire to complete within 2 days of admission. However, previous research by the same team found that the PROM was often not completed, especially for patients who had a shorter length of stay (indicative of shorter survival times). The study sought to understand reasons for this by asking nurses at the service for their views and experiences of using the instrument in this context. A questionnaire with six questions and room for open-ended responses was distributed to 36 nurses at the service and responses were received from 26 (72%).

Nurses' experiences of using and attitudes towards the PROM varied; they reported ways in which the PROM had supported their interactions with patients but also identified some of the difficulties they felt patients experienced in completing the questionnaire. In relation to our theories under test, quotations suggest that a written PROM may make it more comfortable for patients to share information. The authors report that 'Several nurses pointed to its value in identifying symptoms that might not have naturally been discussed without systematic assessment',³⁴² including pain, emotional and existential issues. As one nurse explained, this was because 'it captures problems that patients might write down, but don't speak verbally about'.³⁴² They also indicated that the PROM provided nurses with a structure for their discussions with patients so that they did not forget to ask about things; as one nurse described, 'I become more structured, don't forget important things needed to give good care'.³⁴²

Some nurses described the PROM as supporting the therapeutic process with patients as it gave patients an opportunity to 'highlight his or her problems and distress'.³⁴² They also felt that it was more useful when the PROM was completed in conjunction with a discussion with the patient; as one nurse observed, 'it's a good basis for a discussion about the patient's situation'.³⁴² The authors note that nurses perceived the PROM to be 'a positive complement but not . . . an alternative to other nursing assessments'.³⁴²

However, nurses' concerns about the use of the questionnaire also point to aspects that may constrain a patient's willingness or ability to complete them. The authors report criticisms of the length, wording and format of the questionnaire. Some nurses 'described it as difficult to complete because it was too extensive and time-consuming – "too many questions to fill in adequately"'.³⁴² Nurses also reported 'difficulties for patients in distinguishing between some items, difficulties with the time window related to the past week, as well as questions said to be "too intimate" for some patients and "too intellectual" for others'.³⁴² Some criticisms emphasise particular difficulties in using the questionnaire with palliative care patients. These included practical aspects of the format, 'such as layout that was not tailored to the needs of this patient group, i.e. small print and many pages', and concerns that the questionnaires was too demanding for these patients; one nurse reported that it was 'too comprehensive for all patients to have the strength to complete', and another said that 'patients are too sick, even dying'.³⁴² Linked to this, the nurses responses point to a possible negative effect of completing PROMs for patients: 'tired patients become more tired and irritable'.³⁴²

This study provides some evidence about processes through which completing PROMs might enable patients to share their concerns, and about conditions that might reduce accuracy and openness. It is not clear from this study whether the PROM was completed by patients independently or completed in conjunction with the nurses. We can speculate that a variety of different practices were used by nurses in this setting. This suggests that patients may be more at ease disclosing issues through a written PROM than verbally. Written PROMs can be a more comfortable format for sharing concerns and encouraging

open responses. The findings also suggest that the use of the PROM can also make nurses' enquiries more structured so that they do not forget to ask the patient about certain issues. However, this study raises questions about the extent to which patients' views can be captured by questionnaires that are too demanding for them to complete, such that they are unable to provide complete answers. This 'unintentional inaccuracy' related to capacity to provide full responses may be a particular issue with detailed PROMs, inaccessible formats and patients with little energy. In these contexts, patients may be unable to complete PROMs in a way that adequately represents their views.

This was another small qualitative study in a single setting. The suggested value to patients of being able to write down concerns, and the concerns that patients may struggle to complete the questionnaire, are based on nurses' perceptions. It is unclear how widely these views were shared among nurses in the study, or if they are shared by patients at the palliative care service.

Slater and Freeman³⁴³

This study examined the introduction of the Palliative Outcome Scale (POS) in a day hospice. Patients completed the POS questionnaires at the start of the day in a communal lounge, or took them home to complete. Once forms were collected, staff discussed any identified needs with patients during the day. A focus group was conducted with nine patients who had been using the POS for at least 3 months. Note that the PROMs implementation process for this study does not involve a scheduled consultation meeting, but rather nurses approach patients for discussion if concerns are raised on the POS.

Overall, patients felt that the POS was useful in helping them to communicate their individual needs. The study suggests that PROMs can provide a more comfortable format for patients to share their concerns. This is partly about the privacy of a written format, and partly about feeling more comfortable responding to direct questions.

In relation to privacy, one patient noted that:

It gives you a chance to write something that you don't want the others to hear if you have got some emotional problem with the family or you know and you can write it down knowing that that would be followed up later during the day.

Participant 6³⁴³

A written PROM can allow patients to indicate concerns that they do not want other patients to know about.

Other patient comments point to the importance of format for enabling privacy: the questionnaire used in the study had a section on the front asking about the patient's concerns. This was seen as insufficiently private, with one patient concerned about 'having personal issues written on the front of the form for all to see'.³⁴³

I think it's too open as well for everyone else, for everybody to look at, I think that's one reason I don't fill it in . . . I never fill it in I never have done, it's private.

Participant 8³⁴³

This could be addressed by changing the format, as the authors note: 'The group displayed a high level of consensus to move the section identifying the patient's own problems from the front of the form to the back, which is how it was formatted on the original POS form'.³⁴³

In relation to question wording, patients felt that more direct questions may promote openness. The authors observed that 'The group expressed very strong views that direct questions were needed on the POS forms on whether patients had had suicidal thoughts and feelings'.³⁴³ Two quotations from

patients are given to support this idea that patients may be more likely to share their concerns with direct questions, with a specific suggestion that needing to simply tick a box is more comfortable for patients:

It says have you felt sad, worried or angry it doesn't specifically say outright . . . suicidal, and I think it should be put.

Participant 3³⁴³

It's easier if you've just got to tick the box it's easier than writing it down [you're feeling suicidal] particularly for someone I would imagine who was in deep depression.

Participant 4³⁴³

Another participant also felt that being able to write down their problems was positive and useful:

It makes you think constructively about how you are feeling, and I find putting it into words comforting, just knowing there is someone who is going to read it, and in some cases has the answers.

Participant 7³⁴³

The participants discussed the freedom of being able to express their feelings through the POS form. This may enable patients to adapt to the implications of their illness. The skills of the professionals are paramount in helping patients to reflect on their individual situations at assessment and review. The POS may assist with this process.

However, although patients felt that more direct questions could promote openness, the findings also indicate that some patients found completing the form distressing owing to the phrasing of particular questions. In particular, there were different reactions to the question 'Have you been feeling anxious or worried about your illness or treatment?'. For some, this question provoked offence and distress:

Well I mean when I read that question first I really felt upset and worried and angry I felt like saying well of course I do . . . [this] disease will finish me off what do you think I feel like.

Participant 3³⁴³

Indicating variation in reactions, however, another patient felt that this question allowed them 'to express their frustration':

I get angry and upset because I can't do things, can't do my buttons up it takes me half an hour to do my coat up, I mean there's nothing anybody else can do.

Participant 5³⁴³

These different comments point to the importance of individual context, with different (positive and negative) reactions to the same PROM. The authors note that palliative care patients may adopt either a tolerating or an adapting coping style and that this may influence their engagement with PROMs completion:

Patients using an adapting style may feel more able to confront their problems, while those using a tolerating style may prefer the opportunity to escape from their sick role.

More direct questions may be emotionally beneficial for those adapting, but upsetting for those tolerating. This suggests one aspect of individual context that may affect patient reactions to the PROM and the suitability of direct questions.

Beyond these aspects related to whether or not the PROMs format means that patients feel comfortable about sharing information, the study also indicates that patients may sometimes misunderstand PROM questions. This could mean that the answers they provide do not accurately reflect their symptoms and

concerns. For example, the authors explained that one change to the POS form 'appeared to cause confusion, as participants reported not knowing how to answer these questions'.

A final point to note from this study is that the authors observed that the patient focus group had 'a high level of consensus on physical, psychological, and communication aspects of care with less focus on social and existential aspects of care'. Reasons for the more limited attention to social and existential issues are not explored, but this may suggest that patients find the PROM more useful for identifying some of their needs (physical, psychological) than others.

This was a small study with just nine patients from the same palliative care unit. The authors indicate when views are shared by the group, providing some confidence that the quotations do not just reflect the perspective of one patient. However, additional evidence would help to test out some of the ideas raised by patients, for example looking at whether or not a change in PROMs format affects the information provided.

Although this is a small study, it provides further insight into ways in which PROMs might support patients to give an open account of their concerns and conditions that might affect this. In relation to our focus on patient responses at this stage of PROMs completion, the study suggests that one way PROMs can encourage openness from patients is when a written format provides more privacy. This may be a particularly important consideration in settings where verbal discussions may be overheard: the study took place in a day centre with a communal lounge, and privacy may be less significant in clinical settings with individual patient consultations. The study also points to aspects of context that enable this privacy, specifically a format that means information that patients consider to be private is not on open display.

The study also suggests more direct questions and tick-box formats can support openness from patients, by being more comfortable to answer. However, it points to a potential negative impact of direct questions, with some patients finding these upsetting. The authors' discussion of patients who are 'tolerating' or 'adapting' suggests a useful aspect of individual context that may affect reactions to direct questions and, through this, potentially openness in their answers. The scope for varied reactions to particular questions points to the potential value of more individualised PROMs. A final issue raised by the study relates to patients' ability to understand PROMs. If the format or wording mean that patients are uncertain about what information is needed or how to complete the PROM, the accuracy of answers may be reduced.

***Slater and Freeman*²⁴²**

This paper was based on the same study about the use of the POS in a day hospice discussed above, but it reported findings from a focus group with staff instead of patients. Eight of the nine staff members were included in the focus group (one being unable to attend), and they included four registered nurses, one allied health-care professional and three support staff. Seven of these staff had used the POS since it was introduced in the hospice 9 months previously, and one staff member had used it for 3 months.

Some staff felt that patients did not always complete the form openly. They gave examples of patients who did not express concerns on the form but raised them later in conversations:

It's like when they put low scores on there [the POS], you get them in the bath, giving them a Jacuzzi and they burst into tears and they tell you lots more than comes out on the form.

*Participant 7*²⁴²

This was noted particularly in relation to social issues; another staff member reported that patients might not indicate any difficulties on the form, but then:

later the patient stops me . . . and says . . . I've got this problem and I wonder if you can help me, and it's very rare that a patient will put down any social issues that are . . . bothering them, and this could . . . have a knock-on effect to their general health.

*Participant 6*²⁴²

This suggests limits to the openness promoted through the PROM implied in the patient focus group in the earlier study by Slater and Freeman.³⁴³ As noted in that paper, patients put less emphasis on the role of the PROM in supporting the discussion of social concerns. The focus group with staff suggests that this may be because patients do not record these social aspects on the POS. However, the process behind this is unclear, for example whether patients are reluctant to write down social concerns, or if they feel that recording these issues is unnecessary because they can raise them directly with hospice staff.

Perhaps reflecting their view that patients may admit concerns in person that they do not disclose on the form, some staff felt that talking to patients was more beneficial and more likely to encourage patients to reveal concerns:

I suspect in their position [the patients] you can't beat someone you feel confidence in and you feel comfortable with and all these things come out and no amount of bits of paper is going to change that.

Participant 1²⁴²

One aspect that staff suggested might affect openness was the lack of privacy for patients in a communal setting: the authors noted that:

Staff also had reservations about completing forms for those patients in an open lounge as they thought it affected confidentiality.²⁴²

These concerns about privacy are from the clinicians' perspectives and may not be shared by the patients. However, combined with the concerns about privacy given by patients in the 2004 paper, this suggests that the confidentiality of PROMs feedback can affect openness.

A further area of similarity between the staff and patient focus groups related to the potential for emotional distress as a result of PROMs completion. Staff were concerned that the POS could be upsetting for patients who 'may not wish to confront their problems so directly'. As described by one participant:

For instance . . . you know somebody who has got no energy levels, who's been very tired, very fatigued . . . they might have been trying to hide that, they might not want to come out and might not want to face how things actually are for them.²⁴²

Other staff members said that they wanted the POS to acknowledge the patient's positive achievements and so provide a more realistic balance of the patient's needs and abilities.

As with the 2004 article focusing on patients, these findings are based on a small group in one setting. The article does indicate when views are shared within this group, and the ideas about privacy, patients not raising social concerns and potential distress appear to be held by at least some of the staff and not just a single participant. A further limitation is that staff may not understand or accurately represent patients' attitudes towards the PROM. However, the issues raised by staff are largely in line with the concerns reported in the patient focus group study, including concerns about privacy and the potential for emotional distress if patients do not want to confront their health status.

In relation to the question of whether or not patients complete PROMs accurately, the article suggests that PROMs may be less accurate in capturing social concerns. However, it is unclear if this is about patient reluctance to disclose these issues or some other process. The article also supports the idea that lack of privacy may affect openness, and that completing PROMs may upset patients who are reluctant to acknowledge their deteriorating health.

Hughes et al.³⁴⁴

This study explored health professionals' views of using the POS in routine clinical practice. The authors selected a purposive sample of people who were 'experienced in using outcome measures and the POS' and invited them to take part in a telephone interview. Of this sample, 22 people agreed to be interviewed, but no further details are provided regarding their professional backgrounds or roles. The interviews were not tape recorded; 'verbatim notes' were taken throughout the interviews and analysed using thematic content analysis. The authors describe a number of themes relating to the practical aspects of implementing outcome measures; here we focus on the findings of most relevance to theories under test.

Participants reported that they used outcome measures for a number of reasons, including audit and quality assurance, research and in the care of individual patients, for example during assessments and for the development of care plans. Some participants noted that the POS was difficult to use when patients were too ill and that 'some patients feel that it is very intrusive' or found it 'confusing and upsetting'.³⁴⁴ Some professionals also reported that they found some of the questions 'difficult to ask' or found the responses 'difficult to deal with'.³⁴⁴ In response to this, the authors described how some professionals had altered the questions, ranging from 'vocabulary changes, to making more substantial revisions such as reworking questions or, in some cases, omitting certain questions and including new ones'.³⁴⁴ For example, one participant explained how they had omitted one of the questions to improve its usability in their local setting:

*The question about practical issues was taken out. It confused people and wasn't appropriate in [unit] ... patients no longer hold negative reactions to the POS that they had before in relation to specific questions.*³⁴⁴

The authors acknowledged that 'consideration' of the impact that these changes may have had on the validity and reliability of the POS 'is needed'. Instrument developers usually advise against the rewording or omission of items from standardised PROMs, as this may threaten their psychometric properties.³⁴⁵ This was a small, poor-quality study. Little information was provided about participants' professional background, so it is unclear whose opinions the findings are intended to represent. The interviews were not tape recorded but based on field notes, which leaves open the possibility that interviewers selectively heard or recalled information. They were also conducted by the team who developed the standardised measure (the POS) under investigation in the study. Nonetheless, the authors reported both positive and negative experiences of using the measure. In terms of the theories under test, this study demonstrates that when either staff or patients find PROMs difficult to answer or administer, they adapt the PROM to make it more acceptable and, thus, enable its use in their local circumstances. These adaptations sometimes involved changes or omissions to items, which may have threatened the PROM's psychometric properties.

Hughes et al.³⁴⁶

This paper reports on the patient and staff experiences of implementing the POS in non-specialist palliative care settings in England. A purposive sample of 25 non-specialist palliative care settings was invited to participate in the study, 15 of which agreed to participate. Eight were located in the West Midlands and seven were located in London. Each site was asked to recruit a minimum of 30 patients and complete up to four POS assessments on each patient. However, four sites withdrew from the study before the POS assessments began. Across the 11 participating sites, a total of 21 patients were recruited out of an anticipated minimum of 240. These data alone suggest that the intervention was not well received by participants and the authors set out to explore why. The authors interviewed 13 members of staff and three patients to explore their experiences of using the POS. It is not clear if the interviews were recorded or how they were analysed. The authors' findings largely focus on the difficulties experienced in implementing the intervention; here we focus on those findings of most relevance to our theories under test.

The findings suggest that the participating sites perceived the use of the POS as a research exercise rather than part of their routine clinical practice. The use of the POS was viewed by participants as an 'additional' time-consuming task. However, the authors found that the interpersonal aspects of using the PROM

constrained its use. Some nurses were reluctant to use the POS because they were concerned that it might raise issues they were ill-equipped to manage. For example, one nurse explained, 'I also felt that if I disturbed something whilst I was talking to them, I don't have the psychological back up for them'.³⁴⁶ It is not clear here if the nurse was referring to their own ability to deal with psychological issues or the lack of services to refer on to. Nurses felt that asking patients to complete the POS would be intrusive to the patients' 'personal and quiet time'. They were also worried that asking patients to complete a PROM would disrupt their relationship with patients; for example, one nurse expressed the fear that it would 'tar the relationship a little bit – between the nurses and the patients'.³⁴⁶ Nurses reported that they found it easier to ask patients to complete the PROM if they already knew that patient and had a relationship with them. Finally, nurses also expressed reservations about the validity of the resulting data from the PROM, as they perceived that patients completed the PROM in socially desirable ways to 'please the nurses'.³⁴⁶

This was a small qualitative study with participants who were likely to be among the enthusiasts who engaged in the process of implementing the POS in practice. It is not clear if the interviews were audio-recorded. Although the authors interviewed patients, the findings focus on nurses' views. The authors explained that they used the patient interviews to 'contextualise' the nurses' views. Nevertheless, the study provides an additional layer of evidence for theory testing in relation to the impact of PROMs completion on the interpersonal relationships between patients and nurses. The findings suggest that nurses perceived the use of the standardised PROM as potentially detrimental to these relationships and intrusive to patients and, consequently, were reluctant to use it.

Summary of findings from palliative care

The studies provide useful data to examine how PROMs completion enables or constrains patients' willingness to share issues with clinicians depends on the format of the PROM. Similar to the studies examining the use of PROMs in primary care, patients and clinicians held different views about the value of PROMs in palliative care. Patients saw the use of a PROM as a sign that someone was interested in their feelings and some patients felt that a written questionnaire format made it easier for them to share sensitive topics with clinicians. Other patients questioned the standardised nature of the PROM and felt that it did not always fully capture the complexity and dynamic nature of their problems. Furthermore, some palliative care patients were too ill to complete PROMs. It was also clear that for some patients PROMs completion was an emotional experience, as it forced them to confront issues they preferred to deny. The authors explained these findings in terms of differences in patients' coping strategies: those who preferred to confront their problems may not find PROMs completion problematic, whereas those who preferred to deny their situation would be likely to find PROMs completion difficult. In circumstances in which they wish to continue to ignore these symptoms, patients are more likely to misrepresent their responses on the PROM. As such, PROMs completion may not be a panacea for enabling patients to come to terms with their situation.

In contrast, clinicians in palliative care felt that the standardised PROMs did not fully capture patients' concerns, and perceived that patients were more likely to share their feelings and emotions with them verbally, and often incidentally, while clinicians were caring for patients, rather than through the completion of a PROM. Like the GPs in studies exploring PROMs in the care of patients with depression in primary care,^{80,81,336} clinicians in palliative care settings placed greater emphasis on verbal communication in building rapport and trust with patients. In some instances, nurses were reluctant to use the PROM, as they perceived that it may be detrimental to this relationship. In other instances, nurses changed or omitted the items to make the PROM more acceptable to patients. They felt that PROMs would not support patients to share their concerns unless they had developed this trusting relationship and that the PROM was most useful if used in conjunction with a discussion with the patient.

The next four studies reviewed provide a useful test of the intersection between PROMs format and trust in the clinician. Three studies focus on the use of PROMs by nurses in the initial assessment of patients referred to palliative care services. As such, these studies focus on the start of the process of relationship building between patient and clinician in a context in which the goal of care is maximising patients' quality

of life. Eischens *et al.*³⁴⁷ and Gamlen and Arber³⁴⁸ examine how nurses used standardised PROMs in this context, while Annells and Koch²⁴⁴ explore nurses' views on both standardised and individualised PROMs. In addition, we also review a study that explores the use of an individualised PROM, the SEIQoL-DW, in an oncology setting.²⁴³

***Eischens et al.*³⁴⁷**

This study assessed hospice nurses' experiences and views of using two standardised PROMs in the assessment of patients in a hospice setting. The PROMs in question were the McGill Quality of Life Questionnaire (MQOL) and the Hospice Quality of Life Index (HQLI)-Revised. The MQOL has 18 items covering five domains (physical symptoms, physical well-being, psychological, existential and support) and was specifically designed for use in palliative care settings. The HQLI has 28 items covering three domains (psychophysiological well-being, functional well-being and social/spiritual well-being) and was also specifically designed for use in hospice settings. Eight home care nurses were invited to participate in the study and asked to invite patients to complete the PROMs either independently or with the nurses' help. To be eligible to participate, patients had to be aware of their surroundings, be able to understand English and have the ability to understand and complete the survey. During the study, 37 patients were eligible; 13 completed the MQOL and nine completed the HQLI (total $n = 22$). Nurses were interviewed 1 week after the PROMs were administered to explore their experiences of using the measures and whether they enabled the nurses to become aware of any patient management issues they had not been aware of previously.

The authors reported that nurses felt neither instrument would be appropriate for use on admission to the hospice. For the MQOL, this was because nurses felt that 'a good rapport seemed essential for the patients to answer the questions truthfully'.³⁴⁷ For the HQLI, this was because 'it would make the admission to process seem too overwhelming'.³⁴⁷ The MQOL fared better in terms of its perceived contribution to care, with all nurses reporting that the MQOL had enabled them to identify an area of patient care that they had overlooked. For the HQLI, only one nurse reported it had enabled him or her to identify anything new.

This was a small, poor-quality study. No verbatim quotations are used to illustrate the authors' findings, so it is difficult to assess the extent to which their interpretations are supported by these data. Furthermore, the nurses used the PROMs for only 1 week, which is unlikely to have been long enough for them to fully evaluate the PROMs' contribution to care. Nonetheless, this study provides a first layer of evidence that a good rapport with patients was necessary before standardised PROMs could be used to assess patients, and that their use at this point might have been overwhelming for patients. However, this study did not explore why nurses held these views.

***Gamlen and Arber*³⁴⁸**

This study explored how specialist nurses used a standardised PROM, the Symptoms and Concerns Checklist (SCC), in their first assessments of patients referred to a palliative care service with advanced cancer. An ethnographic approach was used; first assessment encounters between six specialist nurses and patients referred to the service were observed and field-notes were taken. All six nurses were also interviewed to explore their experiences of using the SCC in this context.

The authors observed that nurses placed great importance on hearing the patients' stories in their own words at the first assessment. Not only did this provide them with information on the patient's history and their understanding of the condition, but the words patients used also offered insight into how they were coping with their condition. As one nurse explained in an interview, 'their story and their words tell me what they understand but also what words they use and whether there's anger there or um . . . denial'.³⁴⁸ The priority at the first assessment was to develop rapport with the patient and build trust so that the patient felt comfortable sharing their experiences and feelings with the nurse. A number of nurses in the study felt that the SCC constrained this relationship-building process because it seemed to reduce the

patient's experience to a series of tick boxes, channelled nurses into bombarding patients with questions and was seen as overly bureaucratic. Two quotations from nurses below illustrate these issues:

I like to . . . let them verbalise their concerns rather than handing them a bit of paper and say 'tick boxes', I much prefer to sort of ask . . . rather than bombard them.

Nurse 3³⁴⁸

Some people . . . say 'not another form' and that puts me off straight away . . . this interview is supposed to be about something else . . . it's going on a journey together, exploring . . . where they've come from and where they're at and where they're going . . . It's not as harsh as filling out a form.

Nurse 5³⁴⁸

Consequently, the nurses had found ways to manage the completion of the SCC within first assessments in a way that did not interfere with building this relationship. They did this by delaying the completion of the form until they felt they had established trust and rapport with the patient. Nurse 1 described that she gave the SCC to a patient only after she'd spent considerable time with them, by which time she hoped 'then they feel quite at ease . . . so hopefully by the time I give them that they feel it's all right and . . . miss out what they feel uncomfortable with'.³⁴⁸ This nurse implied that part of this trust building was also giving patients the space *not* to share things if they did not wish to. Nurse 6 also explained, in her interview, 'I never do it straight away', and felt that if the SCC was completed with the patient too early on in their interaction, her assessment became 'disjointed' and she got 'too caught up in going through the questions, "have you got pain?" and all of that'.³⁴⁸ In other words, the SCC could artificially narrow the focus of her assessment.

However, nurses did find the SCC useful to 'validate' and 'consolidate' what they had learned through talking to the patient and felt that it was a useful way of opening up a conversation about specific issues that may require further discussion. Nurse 4 described that the SCC:

[B]rings out a lot of what we've . . . talked about but it kind of consolidates it like 'so you ring [on the tool] pain, so let's talk about pain'.³⁴⁸

They also felt that the SCC prompted them to explore some of the non-physical aspects of the patients' experiences, such as spirituality and sexuality. Nurse 1 explained that the SCC:

[M]akes you actually go there and ask, had you not thought about it, especially with non-physical stuff, which maybe, we sometimes forget.³⁴⁸

This was a well-conducted but small qualitative study of the use of one PROM on first assessment with patients referred to palliative care services. It indicates that if a standardised PROM is used before a certain degree of trust has built up between the patient and the clinician, the relationship-building process can be hampered and the focus of the discussion can be narrowed. This argument has particular resonance in settings in which a trusting relationship between the patient and clinician is particularly important in supporting care and treatment, such as in palliative care and mental health services. The study also suggests that a PROM can be a helpful way of validating information gleaned through discussion with the patient and a useful device for opening up conversation about specific issues that may require further exploration. The study also illustrates that considerable professional skill is required to integrate the use of a PROM into the process of relationship building with patients, suggesting that the PROMs may not be a panacea for improving the communication skills of clinicians. Rather, the effective use of PROMs requires good communication skills to manage their integration into the consultation.

Annells and Koch²⁴⁴

This paper reports the experiences of eight district palliative care nurses during a pilot study to explore the feasibility and acceptability of using PROMs in the first assessment of patients admitted to the service.

The purpose of the pilot study was to test out the use of two different PROMs to enable a team of advisors to decide which PROM to recommend for routine use in the service. The PROMs in question were a standardised PROM, the MQOL, which comprises 18 items covering five domains (physical symptoms, physical well-being, psychological, existential and support) and was specifically designed for use in palliative care settings, and an individualised PROM, the Client Generated Index (CGI), which is an Australian adaptation of the Patient Generated Index.²⁶⁷ With this PROM, the patient selects up to five areas of their life that are most impacted by their condition and rates their severity. To gain an understanding of the importance of each area, patients are invited to 'spend' 12 points on areas of their life they would most like to improve, assuming this was possible.

In this study, 59 palliative care patients were assessed using one of the tools at the first assessment (or soon afterwards). Patients were alternately assigned to be assessed using either the CGI or the MQOL. Data were collected over a 40-week period, during which 71 assessments were made, as some patients were also followed up using these measures. The eight palliative care nurses made written notes of their experiences of using the measures after each assessment and were also interviewed to explore their experiences.

Based on a quantitative analysis of data collected using the tools (the findings are not described in the paper) and a qualitative analysis of the nurses' interviews, the CGI was recommended as the most appropriate tool for use at first assessment. The authors report that the CGI was not likely to be useful for follow-up assessment or patients requiring end-stage terminal care. The rationale for and criteria against which these decisions were made are not described clearly in the paper. However, we can hypothesise that perhaps some of the reasons CGI was deemed not useful for follow-up assessment may relate to the difficulties of using individualised instruments as measures of outcome identified by Farquar *et al.*³³⁴ in terms of the consistency of specifying and then rating cues over time. The authors noted that there were 'therapeutic gains' from the actual interview when administering the CGI, in line with the theory that individualised measures are most useful as a 'conversation opener'. Thus, for the purposes of our synthesis, we focus here on the qualitative findings of nurses' experiences of using this tool during their assessment. The paper focuses on describing the 'intangible' benefits of using the CGI that were not reflected in documented nursing care plans. The authors argue that patients experienced therapeutic gains from completing the measure which extended beyond the remit of what is traditionally considered to be nursing care. The authors describe a number of these 'intangible' benefits in the paper.

The completion of the CGI encouraged patients to reflect on their life in relationship to their current situation. The authors described that the completion of the CGI gave patients 'permission to be emotional' which was interpreted by the nurses as being cathartic for patients. The nurses reported that patients told them it was a relief to discuss their emotions; as one client explained to a nurse, 'I think I needed to have a good cry'.²⁴⁴ The nurses also felt that the completion of the CGI invited clients 'to be open and honest' and to share issues they had not talked about before. When these were raised with carers present, it sometimes resulted in a discussion of these issues between clients and their carer, which was perceived by the nurses as 'therapeutic for client carer relationships'. Nurses also perceived that completing the CGI enabled clients to relate their history to nurses; as one nurse discussed in her interview, 'I feel that the tool really helps the client to tell their story'.²⁴⁴

The nurses observed that the completion of the CGI provided new information to them that they would not have usually uncovered during their assessments. As one nurse explained, 'I wouldn't have picked up on those things normally. They may seem infinitesimal, but they are very real to the client'.²⁴⁴ The nurses perceived that these were issues of real importance to clients; another nurse reported, 'We got to talk about the real stuff. The real stuff that is worrying them'. Thus, completion of the CGI with patients was perceived to enable the sharing of information that would not have been uncovered in the nurses' usual verbal interactions with patients at initial assessment.

This was another small study based on the experiences of a select group of nurses in a single setting. The rationale for the decision-making that the CGI was the tool of choice for the service was not well described. However, our focus was on nurses' views of the CGI in the assessment and care of palliative care patients. These findings suggest that nurses perceived the individualised PROM to support patients to reflect on their situation and give them permission to 'be emotional' and raise issues with clinicians. It appeared that the nurses had integrated the use of the CGI into the process of relationship building with patients, to enable patients to 'tell their story'. However, this study did not explore patients' views of the CGI and, next, we review a study that explored both clinicians' and patients' views of the use of an individualised PROM in an oncology setting.

Kettis-Lindblad et al.²⁴³

This study examined perceptions of a quality of life instrument among patients and oncologists at two hospitals in Sweden. Twenty patients with gastrointestinal cancer completed the SEIQoL-DW and the Disease-Related SEIQoL-DW on a touchscreen computer immediately before their consultation. Patients were then given two copies of the results, one of which was given to the doctor in the consultation. The participating patients were interviewed immediately after each consultation. The eight oncologists who treated at least one participating patient were interviewed at the end of the study. As well as indicating aspects around patient response, the study provides information on whether or not PROM results are used and discussed by clinicians, and we consider these findings in subsequent sections.

In relation to the role of PROMs in helping patients raise their concerns in consultation discussions, the authors suggest that completing PROMs provided an opportunity for reflection, which increased patients' self-awareness:

Several patients acknowledged that the instrument encourages the patient to reflect upon his/her own overall life situation. 'It forces one to think' (Patient 21) and to 'sort the important from the less important' (Patient 6) in the light of the disease.

*Kettis-Lindblad et al. (emphasis in original)*²⁴³

This increased awareness of their priorities and could support patients' ability to raise concerns with clinicians; however, this link from reflection to raising issues is not explicitly stated by the authors.

A more explicit link from completing the PROM to raising issues is discussed in relation to the role of PROMs in making discussion of social and psychological issues legitimate. The authors suggest that completing the PROM gave patients confidence that their feelings mattered: 'simply providing the instrument indicates that a doctor is willing to listen to the patient', and 'patients may be *empowered* by the use of the instrument, since it makes it easier for the patients to *voice their concerns*' (emphasis in original). This is illustrated by the following quotation:

I think it's good in the way that if you have something like this [QOL results], then you can actually tell the doctor that, 'Now I would like to talk about this and that,' and then he'll have to listen.

*Kettis-Lindblad et al.*²⁴³

The article points to aspects of individual context that might affect the value of the PROM in helping patients to raise their broader concerns. Patients felt that the PROM may be more useful for those who are more emotionally distressed, lack a social network or are terminally ill. On the last point, one patient is quoted as indicating that discussion of quality of life issues is more important in later stages of the disease. As one patient explained:

*The worse the stage [the disease] gets . . . the more important [the QOL assessment] gets. Now, when I meet the doctor, we talk a lot about technical matters, how to proceed, which treatments you should get . . . If it turns out that [the disease] gets worse and worse, . . . [the QOL assessment] gets more important.*²⁴³

*Kettis-Lindblad et al.*²⁴³

The nature of the PROM tool may also be a significant aspect of context for enabling these processes of reflection and empowerment. The authors suggest that the relatively open format of the PROM tool used in this study may be important for enabling patient reflection:

*[T]he SEIQoL-DW allows individuals to raise any issues they consider to be important. [...] The focus on areas most important to patients also seems to encourage them to reflect more actively upon their life situations as they relate to their disease.*²⁴³

This was another small study examining patients' and clinicians' views of PROM completion in a single setting. This was a short-term study, lasting from April to August 2004. The participating patients would have had only minimal experience of using the PROM (once during the study), and this might have limited their ability to assess its value. However, the patient interviews provide some evidence in support of the theory that individualised PROMs enable patients to reflect on their situation. There was some evidence to suggest that PROMs completion may 'give patients permission' to raise issues with clinicians. The findings point to possible mechanisms for this process: PROM completion can provide an opportunity to reflect and identify priorities that could then be discussed with clinicians; and providing the PROM may increase patients' confidence to raise issues with clinicians. Aspects of context that may support these processes are also indicated. The patient's individual context and stage in the disease may affect the value of reflecting on quality of life (e.g. this may not be a priority for those who can discuss these issues with their social networks, or for those in earlier stages of the disease who are focused on treatment). In addition, a more open format for the PROM might support individual reflection better than a PROM that asks directly about specific conditions or emotions.

Summary of findings on contextual enablers of and barriers to use of patient-reported outcome measures

These four studies have allowed us to test and refine our theory about the ways in which the format of PROMs may support or constrain the process of relationship building with patients and, consequently, the sharing of information. Gamlen and Arber³⁴⁸ found that the use of a standardised PROM in initial assessments with palliative care patients could threaten the relationship-building process with patients if it is used before this important relationship building has taken place. In contrast, Annells and Koch²⁴⁴ found that the use of an individualised PROM supported this process by offering a means through which patients could tell their story to nurses in their own words and, consequently, share information that really mattered to them. We can hypothesise that because the individualised PROM allowed patients to identify issues in their life that were important to them and did not place restrictions on or standardise the ways in which patients were able to share their story, it more naturally fitted into the ways in which nurses interacted with patients at initial assessments. As Gamlen and Arber³⁴⁸ identified, nurses preferred to let patients share their stories in their own words; Annells and Koch²⁴⁴ found that the CGI enabled patients to do this and, therefore, this supported the relationship-building process. In contrast, the standardised PROM in Gamlen and Arber's³⁴⁸ study, the SCC, narrowed the focus of this information-sharing process and thus threatened or disrupted nurses' attempts to build relationships with patients. Furthermore, Kettis-Linblad²⁴³ also found that the completion of an individualised PROM prompted patients to reflect on their life situation and gave them permission to raise issues with clinicians. However, Annells and Koch²⁴⁴ also noted that the CGI was unlikely to be useful for follow-up assessments. The authors do not explain the rationale behind this judgement, but we can hypothesise that this may relate to the difficulties of using individualised instruments as measures of outcome identified by Farquar *et al.*³³⁴ in terms of the consistency of specifying and then rating cues over time. This supports the theory that individualised measures may have more value as a 'conversation opener' than as an outcome measure.

Chapter summary

In this chapter we tested the overall programme theory (theory 10) that PROMs completion, either alone or with a clinician, enables patients to share or raise issues with clinicians by acting as a reminder, facilitating

self-reflection or giving patients ‘permission’ to raise issues. We also tested the counter-theory (theory 11) that PROMs completion and review may constrain the development of the patient–clinician relationship or hinder the flow of the consultation. We explored how different contextual configurations may shape the mechanisms through which PROMs feedback works and either support or constrain the achievement of the intended outcomes. These included:

- *Theory 12* the structure and format of the PROM and the tension between the length of the PROM and its comprehensiveness and the standardised versus individualised nature of the PROM
- *Theory 13* the existing nature of the patient–clinician relationship, the point in the relationship-building process when the PROM is completed and patients’ and clinicians’ preferences for relationship building
- *Theory 14* if PROMs completion is subjected to monetary rewards and sanctions and whether PROMs data are also intended for use as an indicator of service quality.

To do this, we reviewed studies that compared patients’ views of different PROMs, studies that provided a qualitative analysis of the process of PROMs completion and studies that explored patients’ and clinicians’ experiences of PROMs completion in mental health, palliative care and oncology settings. We now summarise our findings in relation to our theories.

Theory 10

We found evidence to support the theory that PROMs completion prompts patients to engage in a process of self-reflection^{81,243,244,338,343} and enables them to identify and prioritise issues that are important to them.^{243,244,338,339} However, PROMs completion can be an emotional process and the degree to which patients engage with this process may depend on their preferred coping strategy; patients who prefer to deny their current life situation avoid completing PROMs or do not report the true extent of their feelings.³⁴³ We also found evidence to support the theory that completing a PROM signals to the patient that they feel someone is interested in their feelings^{81,328,341,343} and this gives them ‘permission’ to share or raise issues with clinicians.^{243,244,338}

Theory 11

However, we also found evidence to suggest that, under some circumstances, PROMs completion may constrain the development of the patient–clinician relationship or hinder the flow of the consultation. This was dependent on the contextual configurations outlined in theories 12 and 13, which we consider next.

Theory 12

We found evidence to suggest that patients are willing to tolerate a questionnaire that is longer and more complex to complete if it provides a more comprehensive assessment of their health.³²⁸ Some patients felt that the ‘impersonal’ nature of standardised PROMs was helpful in enabling them to share issues with the clinician through PROMs completion.⁸¹ However, others, particularly those with mental health problems, felt that standardised PROMs simply did not capture the complexity or dynamic nature of their symptoms.^{81,340,341} These observations were also shared by clinicians.^{81,242,336,337} Most studies we reviewed did not ask patients or clinicians to directly compare individualised and standardised PROMs; those that did found that patients felt individualised PROMs had greater validity and were less distressing to complete, and clinicians also preferred individualised measures.^{244,329} However, qualitative studies have called into question the validity of individualised instruments as a measure of outcome, due to variations in the ways in which patients and interviewers identify and rate cues over time.³³⁴

In primary care, we found evidence that GPs perceived that standardised PROMs constrained the relationship-building process because they ‘trivialised’ patients’ emotions and resulted in ‘bombarding’ patients with questions in a ‘mechanistic’ way.^{80,336,337} They also found it difficult to incorporate PROMs completion and review into the natural flow of consultations.^{80,336} Similarly, in palliative care settings, we also found evidence to indicate that clinicians perceived standardised PROMs to constrain the relationship-building process with patients during first assessments³⁴⁸ or routine visits.^{344,346} Although not explicitly mentioned by any respondents in the studies we reviewed,³⁴⁸ the difficulties they reported echo

the 'interactional strangeness' of the standardised survey interview, where standardisation is required to support the psychometric validity of the PROM but at the same time restricts opportunities for sense making and relationship building.

Finally, we found some evidence to suggest that, in palliative care and mental health settings, individualised PROMs were perceived as supporting the relationship-building process by enabling the patient to tell their story and prompting opportunities for the patient to self-reflect.^{244,338} However, one study also suggested that individualised PROMs may have more value as a 'conversation opener' than as a measure of outcomes over time.²⁴⁴ Taken together, this evidence suggests that individualised PROMs support the care of individual patients because they allow patients to tell their story in their own words, they prompt patients to self-reflect, patients find them less distressing to complete and clinicians find that they contribute to, rather than detract from, the relationship-building process. However, in this context, they may have more value as a 'conversation opener' between the patient and clinician than as a tool to measure or monitor outcomes over time.

Theory 13

We found that, in primary care and secondary mental health settings and in oncology and palliative care settings, both clinicians and patients felt that having a trusting relationship was necessary to support the sharing of concerns and problems.^{80,81,242,244,336,337,340,341,347,348} Clinicians in primary care, in secondary mental health settings and in oncology placed great emphasis on developing rapport and a trusting relationship with patients through their own verbal interaction with patients, and preferred to let patients 'tell their story' in their own words.^{80,336,337,348} In palliative care, clinicians delayed the introduction of standardised PROMs until this relationship had been built,³⁴⁸ avoided using them³⁴⁶ or omitted or changed items to avoid upsetting patients.³⁴⁴ In secondary mental health settings, patients were reluctant to share their feelings through PROMs completion until this relationship had been developed.^{340,341} This suggests that considerable skill and effective communication skills are required to effectively incorporate standardised PROMs into the flow of the consultation to ensure that trust and rapport between the patient and clinician are not compromised. This indicates that PROMs alone are not a panacea for improving the communication skills of clinicians. In both settings, clinicians held strong views about whether PROMs completion supported or constrained this relationship-building process; this depended on the type of PROM. Patients also expressed preferences for different types of PROM.

Theory 14

We found some evidence to suggest that when patients with mental health problems are worried that PROMs data may be used to restrict access to services, patients may 'game' their responses to PROMs to prevent unwanted outcomes.⁸¹ Parents of children with mental health problems also expressed unease that PROMs data may also be used for political ends to restrict access to or close down services.^{340,341} Some GPs held concerns that patients may inflate their response to PROMs to legitimise sick leave.³³⁷ However, GPs did not feel that the problem of patients 'gaming' their response to PROMs was common.^{81,337} Similarly, in palliative care, nurses expressed concerns that patients completed PROMs in a socially desirably way to please nurses.³⁴⁶

We found evidence to suggest that using PROMs data for a number of purposes and incentivising their use had a much greater impact on the behaviours of clinicians. A case in point was the QOF incentives attached to GPs' use of standardised depression questionnaires in English primary care settings. We have already outlined above that GPs felt that standardised depression questionnaires did not support their management of patients with depression; rather, they constrained the relationship-building process and were difficult to incorporate into the consultation. However, their use was incentivised under the QOF, and if GPs refused to use them they stood to lose financially. Under these circumstances, GPs developed a number of 'workarounds' in the ways in which they administered the standardised questionnaires to mitigate the difficulties they experienced in incorporating them into the consultation.⁸⁰ Some of these strategies threatened the validity of the PROM as a measure of depression severity, thus undermining its use in the care of individual patients.⁸⁰ Furthermore, they also used other tactics to circumvent being penalised for avoiding the use of standardised depression

questionnaires, such as not coding patients as having depression in the first place, thereby rendering the QOF indicators an inaccurate representation of the practice's care of people with depression.⁸¹ This provides an example of an instance in which attempts to use the same PROM for multiple purposes and incentivising their use can undermine their value in serving both purposes. In this case, the key problem was that GPs did not perceive standardised depression questionnaires to be helpful in the care of individual patients.

In palliative care, where the use of PROMs was not incentivised, nurses largely avoided use of the PROMs when they were concerned that they might threaten the patient–clinician relationship.³⁴⁶ However, we also found evidence to suggest that those who tried to implement standardised PROMs also engaged in adapting the standardised PROM to make it more acceptable to patients. These adaptations involved changing or omitting items, which might also have compromised the psychometric properties of the PROM.

None of the studies reviewed in this chapter explored what happened in clinician–patient interactions during or following PROMs completion, or whether patients actually did raise issues during the consultation following PROMs completion. Furthermore, the value of PROMs in improving patient care is also likely to depend on whether or not and how these issues are discussed within the consultation and what action is taken as a result. It is these issues that we turn to in the next chapter.

Chapter 9 Patient-reported outcome measures as a tool for raising clinicians' awareness of patients' concerns

Introduction

In this chapter, we focus on reviewing the theory that PROMs act as a tool for raising clinicians' awareness of patient concerns, which will then be discussed during the consultation and acted on. The clinician–patient consultation has been much researched and remains deeply contested, in terms of both the power struggles that operate therein and the ways in which these have been conceptualised in the academic literature.^{302,308,349} It is a time-pressured environment, raising questions about what is discussed, what gets left out and how this is negotiated between the clinician, patient and, often, the patient's carer. The feedback of PROMs data during the consultation is expected to offer a vehicle through which patients' concerns can be more effectively communicated to clinicians, who will raise these issues with patients and thus give primacy to the patient's agenda during consultations.

In this chapter, we build on and extend a theory-driven review of PROMs feedback previously conducted by a member of the review team.⁶² Systematic reviews of PROMs feedback in the care of individual patients with long-term health conditions suggest that PROMs feedback leads to increased discussion of problems during the consultation, and helps clinicians to identify and detect patient problems, but has less impact on patient management or patient outcomes.^{42,43,48,49,53,54} For example, Chen *et al.*⁴³ found 'strong' evidence that the feedback of PROMs data improves patient–clinician communication and 'some' evidence that it improves the monitoring of treatment response and the detection of patients' problems. However, they found 'weak but positive evidence' that PROMs feedback leads to changes in patient management and 'a great degree of uncertainty' regarding whether or not PROMs feedback improves patient outcomes. Chen *et al.*⁴³ suggested that greater impact of PROMs feedback may be found where PROMs are fed back for a sustained period of time to multiple stakeholders with feedback that is clear and easy to understand and sufficient training for health professionals. Kotronoulas *et al.*⁴² found significant increases in the frequency of discussions 'pertinent to patient outcomes' but little impact on referrals or clinical actions in response to PROMs data. This suggests there may be a 'blockage' between the identification and discussion of the issues raised by PROMs and the ways in which clinicians respond to these issues. In this chapter, we focus on understanding the flows and blockages along this implementation chain and the contextual configurations that may support or constrain how each step leads to the achievement (or not) of the next.

In this pathway, it is assumed that the patient completes a PROM and the individual patient's PROM scores are fed back to the clinician prior to or during the consultation. It is hypothesised that this will then alert the clinician to problems or issues that are of concern to the patient. These scores could be at a single point in time but also denote changes in PROMs scores over time, to enable the clinician to monitor the impact of treatment on the patient's health. It is then assumed that this increased awareness will prompt the clinician to further explore and discuss any problems identified with the patient and, subsequently, take action to address them, through, for example, referring to additional services, giving advice, changing current treatment or suggesting other treatment options. This is then assumed to lead to improved patient outcomes. We start by outlining the specific programme theories to be tested.

Candidate programme theories

We draw on the process of theory elicitation undertaken in *Chapter 7* to identify some provisional theories that seek to explain how PROMs feedback is expected to work. It is important to note here that, initially, these theories are provisional and represent the simple assumptions regarding the circumstances in which and processes through which PROMs feedback is expected to lead to intended outcomes. The purpose of synthesis is then to test these theories in relation to the literature and refine them to produce a more sophisticated understanding of how PROMs feedback works in practice. The goal is to *probe* these theories rather than *prove* that they are right or wrong. Here we outline some initial candidate theories for testing in our evidence review and draw on Street²⁸⁹ and Santana and Feeney's²⁸⁷ models of the processes through which PROMs feedback is expected to improve communication, patient management and patient well-being during the consultation.

One theory is that PROMs feedback leads to increased discussion during the consultation and this in itself has a direct impact on improving patient well-being.^{287,288} Therefore, we start by testing theory 15, that PROMs feedback results in the increased discussion of HRQoL issues in the consultation and this promotes increased patient well-being.

Another theory (theory 16: PROMs feedback increases the discussion of HRQoL issues during the consultation, which in turn influences patient management and improves patients' well-being) is that PROMs may improve communication and also improve patient well-being indirectly, through increased agreement between the clinician and patient about the goals of treatment, and changes to patient management.^{287,289}

However, a counter-theory is that PROMs feedback may not change clinicians' communication practices during the consultation. For example, observational studies of doctor–patient communication in oncology settings indicate that doctors focus more on medical/technical issues, with psychosocial issues discussed only briefly.^{64,350} PROMs feedback may not be sufficient to shift clinicians' communication practices during the consultation and so the focus of the consultation may remain on medical issues. Therefore, we test the counter-theory, theory 17, that PROMs feedback does not change clinicians' communication practices during the consultation.

We now test and refine these theories.

Evidence review

To test these theories, we use oncology as a case study. It is in this area that much of the work has been done to explore whether or not PROMs feedback leads to increased discussion of patients' problems during the consultation, and it has been the focus of several systematic reviews.^{42,43,59} We start by reviewing evidence from a purposive sample of RCTs of PROMs feedback in oncology. The majority of those selected have assessed doctor–patient communication directly through audio recordings of the consultation. Some have also measured one or more of the other intended outcomes (changes to patient management and patient well-being). The trials reviewed below were selected on the basis that they were most relevant to testing and refining the theories outlined above.

Evaluating the impact of PROMs feedback as part of a RCT presents a number of methodological challenges, as the intervention aims to change the behaviour of both patients and clinicians. Many trials have randomised patients as the unit of analysis. This approach means that clinicians see both control and intervention patients, increasing the risk of contamination effects, which may dilute the measurement of the overall intervention effect. Alternatively, when clinicians are the unit of randomisation, patients seen by the same clinician are more similar to each other than to patients seen by other clinicians, and this 'intercluster correlation' must be addressed in the analyses; however, few trials have done this.³⁵¹ These caveats need to be taken into account when interpreting the findings of the trials reviewed below.

Velikova et al.³²⁷

This RCT examined whether or not regular, repeated feedback of HRQoL information to clinicians would result in increased discussion of HRQoL during the consultation and would improve patients' well-being. Patients ($n = 286$) who were about to start treatment for cancer in one specialist centre were randomised to one of three arms: (1) intervention – regular completion of two PROMs on a touchscreen computer prior to the consultation, and the EORTC QLQ-C30 and the Hospital Anxiety and Depression Scale plus feedback to clinicians during the consultation; (2) attention control – regular completion of the two PROMs but no feedback to the clinician; and (3) control – no completion of the PROMs and no feedback. Patients in the control and intervention arms completed the PROMs over three visits. All 28 oncologists who worked in the unit participated in the study. They received print-outs of the patients' current scores and changes over time as line graphs for each subscale score and received training in the interpretation of the scores.

The primary outcome for the trial was the disease-specific PROM, the FACT-G, which has four subscales: physical well-being, social or family well-being, emotional well-being and functional well-being. All consultations were audio recorded, and a content analysis was performed to identify whether or not issues in the EORTC QLQ-C30 were discussed and, if so, the number of issues discussed. Medical actions (decision on cancer treatment, symptom or supportive treatment, investigations and referrals) and non-medical actions (advice on lifestyle, coping and reassurance) were also recorded. Communication and actions were compared during the third consultation and outcomes were compared at the end of the trial.

The authors found that patients in the intervention arm of the trial had significantly greater improvements in well-being (total FACT-G score) than those in the control arm but not those in the attention control arm. The attention control arm also had statistically significantly higher FACT-G scores than the control arm. The number of EORTC QLQ-C30 symptoms discussed (mean 3.3) was greater in the intervention arm than the control arm (mean 2.7), but there were no statistically significant differences in the number of non-specific or functional issues discussed between the control and intervention patients. Clinicians explicitly referred to HRQoL data in 66 out of the 103 (64%) intervention encounters. Subgroup analyses revealed that improvements in patient well-being were associated with explicit use of HRQoL data in the consultation. There were no differences in either medical decisions or non-medical actions between the two arms.

Clinicians perceived that HRQoL data were most useful for providing an overall assessment of the patient (69%) but that they were less useful for providing additional information (33%) or identifying problems for discussion (27%). Clinicians reported that HRQoL data contributed to patient management in 11% of consultations. They reported that they did not use HRQoL information if it was 'irrelevant for the purpose of the encounter or irrelevant to patient's major problems'.³²⁷

This trial found that PROMs feedback led to an increase in the number of symptoms discussed in the consultation and an improvement in patient well-being that was associated with specific mention of HRQoL issues. In terms of the theories under test, these findings suggest that the increased discussion of symptoms may lead directly to improvements in patient well-being, irrespective of whether or not any changes to patient management occur. The authors hypothesise that this may have been especially helpful in their centre, where patients saw different clinicians sequentially, suggesting that PROMs completion may support information transfer between clinicians. However, the trial also found that PROMs feedback did not lead to changes in the number of functional issues discussed, which suggests that the consultation remained predominantly focused on medical issues.

Furthermore, patients in the attention control arm also experienced an improvement in well-being, suggesting that PROMs completion alone may be beneficial to patients, although it appears that this did not lead to more symptoms being discussed during the consultation. This finding may be due to a 'contamination effect', as clinicians in the attention control arm were also exposed to intervention patients and this might have also changed the clinicians' behaviour with patients in the attention control arm.

The findings also hint that clinicians do not see HRQoL information as providing 'new' information about the patient and may not use the information if they do not see it as a relevant to the purpose of the consultation. The patients in this study were those receiving treatment for cancer, where the focus is on curing or limiting the spread of the disease, rather than palliative patients. In this setting, the clinician's priority may be maximising quantity of life provided that side effects are tolerated by patients.

Velikova *et al.*³⁵²

This paper reports on a secondary analysis of the three-arm trial reported by Velikova *et al.*³²⁷ to explore the mechanism through which increased discussion of symptoms during the consultation leads to improvements in patients' well-being. They tested the hypothesis that PROMs feedback improves well-being because it 'ensures a continuous flow of subjective symptoms/functioning information from the patient to the medical team' and improves communication about non-medical problems. The team developed a new questionnaire, the Medical Care Questionnaire (MCQ), to measure patients' perceptions of the continuity and co-ordination of medical care. The questionnaire was used in the trial but, at that time, had not been validated. The team subsequently assessed the psychometric properties of the MCQ and published the findings from the trial using the MCQ. However, part of the validation process for this instrument included data from this trial. Thus, the trial data were used to confirm the psychometric properties of the instrument and at the same time the instrument was used to explain the findings, which appears a somewhat circular argument.

This instrument consisted of three subscales:

1. communication – measures the extent to which the patient feels comfortable discussing non-medical issues
2. co-ordination – measures the extent to which patients perceive that there is continuity and co-ordination of information
3. patient preferences – measures the extent to which patients prefer to see their usual doctor.

The authors also report findings from end-of-study questionnaires completed by patients and doctors that explored the perceived usefulness of the intervention.

The authors found that patients in the intervention arm rated communication with clinicians (on the MCQ) statistically significantly higher than patients in the control arm, but not differently from patients in the attention control arm (in which the patients completed the PROMs but the scores were not fed back to clinicians). There were no differences between the three arms in patients' perceptions of the co-ordination of care or their preferences for seeing the same doctor.

In the intervention arm, 86% of patients perceived that questionnaires were useful for telling doctors how they were feeling, compared with 29% in the attention control arm (where no feedback occurred). More patients in the intervention arm than in the attention control arm felt that the doctor considered daily activities (65% vs. 53%), emotions (87% vs. 71%) and quality of life (90% vs. 74%). However, it is not clear if these differences were statistically significant. The authors concluded that PROMs feedback increased patient well-being because it focused the 'consultation on topics important to patient and facilitated discussion of non-medical issues'.³⁵²

In terms of the theories under test, this study suggests that PROMs feedback to clinicians may increase patient well-being because it results in patients feeling more comfortable discussing non-medical issues with clinicians. However, it is not clear from the findings of this trial whether this is a function of PROMs completion alone or PROMs completion and feedback to clinicians. Recall from Velikova *et al.*³²⁷ that patients who completed the PROM with no feedback of scores to clinicians also experienced statistically significant improvements in well-being, compared with those in the control arm. In this paper, there were no statistically significant differences in MCQ communication scores between the intervention and intervention control arms and the authors do not report whether or not the MCQ scores for

communication in the attention control arm were statistically significantly different from those in the control arm. Thus, improvements in well-being and feeling comfortable discussing non-medical issues could occur because patients completed a PROM and this made the patient feel they had permission to raise issues with clinicians, rather than occurring as a result of any changes in clinicians' communication practices.

However, the findings from the end-of-study questionnaire provide some evidence that, compared with attention control patients, patients in the intervention group perceived clinicians to be more attentive to their concerns. This suggests that some changes in clinicians' communication practices did occur in the intervention group compared with the attention control group.

Takeuchi et al.⁶³

This paper reports on a more detailed content analysis of all audio-recorded consultations collected during Velikova *et al.*'s³²⁷ three-arm RCT, to further explore the dynamics of doctor–patient interaction during these consultations. The authors' analysis aimed to explore whether or not the differences in the number of HRQoL issues discussed between the control and intervention groups were maintained over time. They also sought to specifically test the theory that PROMs feedback would prompt oncologists to initiate discussions about problems reported in the PROMs. Their analysis also sought to test the hypothesis that the severity of a patient's problems would be related to whether or not the problem was discussed.

Data were from 198 patients who completed four consecutive consultations with one of 28 oncologists and, in total, 792 consultations were included in the analysis. The data were analysed using the same study-specific checklist as Velikova *et al.*,³²⁷ which assessed whether or not issues in the EORTC QLQ-C30 and Hospital Anxiety and Depression Scale were discussed, and whether or not other non-specific symptoms were discussed. In addition, the authors noted whether clinicians or patients initiated discussion about the issue.

The authors found that the largest impact on increased discussion of symptoms in the intervention group occurred following the first feedback of PROMs to clinicians. There was a reduction in the number of symptoms discussed between the first and third consultations for all groups. However, the significant differences in the number of symptoms discussed between the intervention and control group and between the intervention and attention control group were maintained over time. The authors hypothesise that this finding may be because at the first consultation, the PROMs feedback provided 'new' information to clinicians and, thus, was more likely to be discussed. There were no statistically significant differences in the number of functional issues discussed over time between the three groups.

During this first consultation, for all arms of the study, discussion of symptoms and functional problems was more likely to be initiated by the patient, with the exception of dyspnoea and bowel problems, which were more likely to be discussed by the clinician. Discussion about symptoms were more common (26–64%) than discussion of functions (4–41%). Overall, there was no impact of study arm on whether or not the discussion of symptoms or function was initiated by the clinician. In terms of the theories under test, this finding indicates that PROMs feedback did not make it more likely that clinicians would raise issues with patients. Although discussion increased, this was largely through the patient, rather than the clinician, initiating a discussion. They also found that the severity of patients' symptoms was predictive of whether they were discussed in consultations: when patients had more severe symptoms, these were more likely to be discussed. However, there was no relationship between the severity of patients' functional problems and depression and discussion; thus, patients with more severe functional or emotional problems were no more likely to have these discussed than those who had less severe functional problems. This suggests that PROMs feedback did not change the ways in which clinicians discussed functional problems.

In terms of our theories under test, this study suggests that PROMs feedback did not substantially change clinicians' communication practices during the consultation. Although PROMs feedback led to increases in the number of symptoms discussed and patients with more severe symptoms were more likely to have them discussed, for most symptoms patients were more likely to initiate the discussion. The authors

conclude that the PROMs feedback 'did not . . . influence the standard medical structure, where patients are encouraged to report their own experience'.⁶³ They hypothesise that the presence of the PROM feedback may have served as a prompt for both patients and oncologists. Furthermore, they also found that symptoms were more commonly discussed than functional problems and that discussion was related to severity of symptoms but not functional problems. The authors interpret these findings as indicating that patients did not feel that oncology consultations were 'an appropriate forum to raise these issues'.⁶³ We would also hypothesise that these findings indicate that clinicians did not feel it was within their remit to discuss these issues. In summary, the authors conclude that PROM 'feedback alone does not overcome existing barriers that prevent oncologists from discussing functional issues with their patients'.⁶³ The authors consider that training oncologists in communication skills and providing guideline-driven recommendations for managing symptoms and functional issues may address these problems.

Mills et al.³⁵³

The hypotheses that guided this trial are left largely implicit by the authors, who indicate that the trial aimed to examine whether or not 'regular recording of QOL [quality of life] data by patients would have beneficial outcomes'. The implicit assumption underlying this trial appeared to be the idea that if patients with inoperable lung cancer regularly completed a quality-of-life diary, they would share the information in the diary with the different clinicians (e.g. nurses, oncologists, GPs) caring for them; the clinicians would address their concerns and the patient's HRQoL would improve. Patients were the unit of randomisation. The intervention consisted of patients completing a weekly diary that comprised the EORTC QLQ-C30 and the associated lung cancer-specific questionnaire: the Quality of Life Questionnaire LC13. They were asked to complete this diary for 16 weeks. Patients kept the diary at home and were asked to 'share the information with any health professional involved in their care'. Clinicians in hospital received a training session about the diary but primary care clinicians received this information via e-mail. There was no formal feedback of the diary information to clinicians. Control group patients did not complete the diary.

Patients included in the study had inoperable lung cancer; of 498 screened, 266 met the inclusion criteria and 115 (43%) agreed to participate. The reasons for not participating included being too distressed (32%) or too unwell (19%) or undisclosed 'other' reasons (36%). By month 4, 57 of these patients remained in the trial (50%). The most common reasons for attrition were related to the intervention itself: 58 patients did not complete the diaries fully, 15 withdrew and 15 did not complete the diary at all. Nineteen patients did complete the diary. These data suggest that the intervention was not initially appealing to patients and that patients found it difficult to stay engaged with it.

The main outcome for the trial was the patient's HRQoL as measured by the Functional Assessment of Cancer Therapy – Lung (FACT-L), a disease-specific version of the FACT-G which contains four subscales as described in the discussion of Velikova *et al.*³²⁷ plus a 7-item lung cancer-specific subscale. Changes in scores over time were used to compare control and intervention groups. The primary outcome was the Trial Outcome Index, which is the sum of the Physical Well-Being, Functional Well-Being and 'additional concerns' subscales of the FACT-L. They also measured quality of life using the Palliative Care Quality of Life Index. Outcomes were measured at baseline, 2 months and 4 months. Communication was measured by patient self-report using a modified checklist from Velikova *et al.*,³²⁷ in which patients were asked whether, over the last 2 months, any health professional had asked them about a list of issues. Patients were also asked if they had shared information in the diary with any health professionals.

The authors found no significant differences in the Trial Outcome Index between the intervention and control groups. However, patients in the intervention group experienced a significant deterioration in their total FACT-L scores and total FACT-G scores, compared with the control group. These differences remained when possible confounding factors were controlled for. This suggests that the intervention has a deleterious effect on patients' HRQoL. Of the 57 patients in the intervention group, 13 (23%) indicated that they had shared information in the diary with any clinician. There were no significant differences between control and intervention group patients in terms of whether or not any of the HRQoL issues were discussed with clinicians.

In terms of the theories under test, this study suggests that, for patients with incurable disease, frequent (weekly) regular completion of a PROM over a sustained period without formal feedback to clinicians has a deleterious impact on their HRQoL. It also suggests that patients rarely share HRQoL information with clinicians when they complete a PROM if there is no formal feedback to clinicians. The authors identified a number of reasons for the findings. First, they could not rule out the possibility that patients in the intervention arm had worse disease progression than patients in the control arm, because data on disease progression were not available to the research team. However, there were no baseline differences between patients in the two groups, which suggests that alternative explanations were also required.

The authors argued that, as no formal feedback was provided to clinicians, this limited the clinicians' ability to raise or address any concerns that may have been apparent in diary data. We can also hypothesise that if patients knew that there was no formal feedback mechanism to clinicians, this may have sent them the message that clinicians were 'not interested' in their HRQoL and, consequently, patients may have been reluctant to raise issues with clinicians. It may be that it is the completion of the PROM plus the expectation of feedback to clinicians that gives patients 'permission' to raise issues with clinicians.

The authors highlight that regular completion of a diary in this group of patients with incurable disease may also have led to patients dwelling on their disease progression and ruminating on their disease. The diary would have enabled patients to see any deterioration in their symptoms, which may have been emotionally distressing. Furthermore, the emotional impact was likely to have been exacerbated as, unlike for patients undergoing curative cancer treatment, there was no prospect that their condition would improve or was amenable to treatment.

Finally, the authors highlight that completing a HRQoL instrument may have raised patients' expectations that clinicians could 'control or influence' their HRQoL, and, when these expectations were not met, this resulted in a negative impact on patients. We can suggest a subtly different hypothesis, that completion of the HRQoL instrument may have raised patients' expectations that these issues would have been noticed or discussed by clinicians and, when they were not, this was emotionally distressing for patients.

Detmar et al.³²⁶

This randomised crossover trial examined whether or not the regular feedback of HRQoL data to doctors would increase the discussion of HRQoL issues during the consultation. The trial took place in one specialist cancer treatment centre in Amsterdam. In this trial, clinicians ($n = 10$) were the unit of randomisation. Five doctors were randomly allocated to an intervention group and five were allocated to a control group. Each doctor was asked to recruit a minimum of 10 patients to the trial. A washout period of 2 months followed, after which clinicians in the control arm were switched to the intervention arm and clinicians in the intervention arm were switched to the control arm. Each doctor then recruited a further 10 patients. Patients were receiving palliative chemotherapy. The intervention group patients were asked to complete the EORTC QLQ-C30 questionnaire before each visit to the doctor and the scores were fed back to both the patient and the doctor, along with bar charts showing patients' current and previous scores on each subscale.

The primary outcome of the trial was the discussion of HRQoL issues during the consultation. All consultations were audio recorded and analysed using a checklist that determined whether issues in the EORTC QLQ-C30 were discussed and summed the total number of issues discussed in the consultation. Communication at the fourth patient visit was compared between intervention and control arm patients. The total number of actions taken during each consultation was also calculated as an indicator of impact on patient management. Patient satisfaction was measured using a 5-item satisfaction questionnaire, and patient baseline and end-of-study HRQoL was measured using a generic measure; the SF-36.

The authors found statistically significant differences in the total number of HRQoL issues discussed in the fourth consultation between the intervention (mean 4.5) and control (mean 3.7) arms. Social functioning, fatigue and dyspnoea were discussed statistically significantly more frequently in the intervention group than in the control group. There were hardly any management actions taken in either arm and differences

were not statistically significant. There were no statistically significant differences in overall satisfaction between the two arms, but patients in the intervention arm scored higher on satisfaction with emotional well-being. There were no statistically significant differences in end-of-study quality of life between the control and intervention arms, but patients in the intervention arm showed greater improvement in mental health scores and role function than those in the control group (two out of eight subscales on the SF-36). As the authors acknowledge, caution is needed in interpreting the latter findings due to multiple testing, which increases the risk of false-positive results. Almost all patients (97%) in the intervention arm felt that the EORTC QLQ-C30 provided an accurate reflection of their HRQoL and 57% reported that it had been explicitly mentioned during the consultation. All doctors indicated that the summary profile had provided a useful overall impression of patients' symptoms and functional health and that it had facilitated communication, particularly regarding psychosocial topics.

The authors acknowledged that this trial had a number of limitations. They found evidence that clinicians initially allocated to the intervention changed their behaviour towards control group patients when they crossed over, as they were more likely to discuss emotional issues at baseline. Furthermore, in this trial, clinicians were the unit of analysis. Patients seen by the same clinician are more similar to each other than to patients seen by other clinicians. However, the authors did not take account of this 'intercluster correlation' in their analyses. In terms of the theories under test, this study suggests that PROMs feedback does lead to an increase in the number of HRQoL issues discussed in the consultation, but this does not necessarily lead to changes in patient management or improvements in patient well-being. This may be because the authors used a generic measure of HRQoL, the SF-36, which did not reflect changes in the disease-specific aspects of HRQoL that may have been particularly improved by their discussion during the consultation. It was also not designed to detect the perhaps transient improvements in well-being that may occur if patients feel listened to in the consultation. It may be that the single patient satisfaction item covering satisfaction with emotional well-being reflected this. However, it could also mean that changes in patient management are necessary to bring about changes to a patient's well-being or HRQoL.

Berry et al.³⁵⁴

This RCT aimed to examine whether or not the feedback of HRQoL to clinicians increases the discussion of HRQoL issues during the consultation. This trial was conducted in two specialist cancer treatment centres in the USA. The participants were cancer patients starting a new treatment regimen. Of the 1104 patients assessed for eligibility, 660 entered the trial, and 262 of the 295 clinicians who would potentially come into contact with the patients agreed to take part. These included oncologists and nurses. However, it is not clear exactly how many of these clinicians actually came into contact with patients. Of the 262 clinicians who agreed to participate, 113 returned a post-intervention questionnaire; of these, 57 reported receiving at least one report.

Patients in both the control and the intervention groups completed the PROM (Electronic Self-Report Assessment of Cancer) on a touchscreen computer in the clinic prior to the appointment. The PROM was completed once before starting treatment and a second time 4–6 weeks later, at which point treatment-related symptoms and quality of life issues (SQLIs) were expected. In the intervention group ($n = 327$), at this second time of PROMs completion, the summary report was printed and given to the clinician or attached to the patient's chart immediately before the appointment. Each SQLI at or above a threshold level was indicated by colour and bar graph height. The report did not provide recommendations for addressing SQLIs. In the control group ($n = 333$), the clinicians did not receive the summary report. The consultations at this second stage were recorded to assess discussion of SQLI. The recordings were analysed to assess whether or not reported SQLIs were discussed. Communication outcomes were calculated as whether or not the symptom was discussed and who initiated the discussion (the patient or the clinician). The study also compared visit duration between the two groups.

The authors found that whether or not the symptom was more likely to be discussed in the intervention group depended on whether or not the problem was reported 'at threshold'. When all 18 symptoms and quality of life issues were considered together as a whole, when the symptom was reported at threshold,

the odds of it being discussed were nearly 29% higher in the intervention group than the control group, a difference that was statistically significant. In 75 out of the 295 recorded consultations in the intervention arm (25%) clinicians made direct reference to the Electronic Self-Report Assessment of Cancer report. The authors then examined whether each of the 18 symptoms and quality of life issues were more likely to be discussed in the intervention or control arm, when reported at threshold. The authors found that some symptoms were commonly discussed in both arms, irrespective of whether they were at threshold. These included symptoms that were regulated by accreditation bodies (e.g. pain) or those that were likely to be affected by supportive care medications (e.g. nausea). In terms of the theories under test, this suggests that PROMs did not change whether or not clinicians enquired about these issues.

Two issues, sexual functioning and social functioning, were more likely to be discussed when they were reported at threshold in the intervention arm. In terms of the theories under test, Berry *et al.*³⁵⁴ suggest that, in these instances, the Electronic Self-Report Assessment of Cancer report may have acted as an 'ice-breaker', giving patients or clinicians 'permission' to discuss sensitive issues. However, relative to the number of times it was reported at threshold, the number of times sexual function was actually discussed in the consultation, even in the intervention group, was relatively very small. In the intervention group, impact on sex was reported at threshold 122 times, but the clinician initiated discussion about it 20 times, approximately 1 out of every 6 times it was reported as a problem. Compare this with pain frequency, which was reported at threshold 64 times but the clinician initiated a discussion about it 155 times. Similarly, nausea was reported at threshold 21 times and yet clinicians initiated a discussion about it 132 times.

Furthermore, two other issues, cognitive functioning and concentration, were not more likely to be discussed in the intervention arm when they were reported at threshold. In terms of the theories under test, this finding would suggest that clinicians actively avoided the discussion of these issues, even when the Electronic Self-Report Assessment of Cancer indicated that they were a problem. Berry *et al.*³⁵⁴ hypothesise that this was because there is 'little evidence for how to avoid or treat cognitive dysfunction without diminishing dose during cancer therapy', suggesting that priority was given to the curative function of chemotherapy. Berry *et al.*³⁵⁴ also theorise that 'clinicians may choose to focus on SQLIs that they can make a difference' to; in other words, clinicians preferred to discuss things that they were able to do something about and avoided discussing issues that they could not treat, even when these were problematic for patients. Thus, in summary, this study suggests that PROMs feedback does not change clinicians' communication practices during the consultation. They continue to focus on medical issues, whether or not these are actually a problem for the patient. Although PROMs may act as an icebreaker for sensitive issues such as sexual function, these issues were still discussed relatively rarely.

Summary

These trials represent a purposive sample of all the trials examining the impact of PROMs feedback in oncology; trials were selected that were most relevant to testing and refining our theories. Although we only reviewed four trials (reported across six studies) in detail, between them, they produced a complex pattern of outcomes that require considerable analytic reflection to explain. Here we interpret the findings of these trials in relation to our theories in order to test and refine them and thus explain this pattern of outcomes.

Theories 15 and 16

We found some evidence in support of theory 15. Velikova *et al.*³²⁷ and Detmar *et al.*³²⁶ found that PROMs feedback does increase the discussion of symptoms during the consultation, but only Velikova *et al.*³²⁷ found that this leads directly to improvements in patient well-being. A longitudinal analysis of all consultations in Velikova *et al.*'s trial found that this increase in discussion of symptoms was sustained over time.⁶³ An analysis of secondary outcomes of this trial³⁵² suggested that the mechanism through which PROMs feedback leads to improvements in well-being is patients feeling comfortable discussing non-medical issues with clinicians. Findings from the secondary analysis of this trial suggest that more patients in the intervention arm than in the attention control arm felt that their clinician was attentive to their daily activities, emotions and quality of life. This indicates that PROMs feedback led to more changes in clinicians' communication behaviours towards intervention patients than in their behaviour towards attention control arm patients. Improvements

in well-being in the intervention group were associated with explicit reference to the HRQoL data, further supporting the idea that changes in clinicians' behaviour contributed to increases in patients' well-being.

We did not find any evidence to support theory 16, that PROMs feedback may indirectly improve patient well-being by improving communication, which in turn leads to improvements in patient management. Detmar *et al.*³²⁶ found that PROMs completion increased the number of HRQoL issues discussed in the consultation but this did not lead to changes in patient management or to widespread improvements in well-being; only two out of eight subscales on the SF-36 showed improvement. These findings may have occurred because the SF-36 was not responsive enough to detect the more subtle changes in patients' well-being that occur as a result of feeling listened to by a clinician. They may also indicate that management changes are required to improve patients' well-being.

However, we also found some evidence to support a new counter-theory, that PROMs completion may reduce patient well-being. Mills *et al.*³⁵³ found that when patients with inoperable lung cancer frequently completed the same PROM on a regular basis, but the results were not fed back to clinicians, few patients shared information with clinicians and patients experienced a deterioration in their quality of life. These findings could have occurred because this group of patients found frequent completion of a HRQoL instrument distressing and increased rumination about their condition. The findings may also have occurred because the lack of feedback meant that clinicians were limited in their ability to raise or address issues with patients. Furthermore, HRQoL instrument completion may have raised patients' expectations that clinicians were interested in their HRQoL but the lack of discussion of these issues by clinicians had a negative impact on patients. It may be that it is the completion of the PROM plus the expectation of feedback to clinicians that gives patients 'permission' to raise issues with clinicians and enables them to feel that the clinician is interested in their HRQoL.

Theory 17

We also found some evidence to support the theory that PROMs feedback does not substantially change clinicians' communication practices during the consultation. Detmar *et al.*,⁶⁴ Velikova *et al.*³²⁷ and Takeuchi *et al.*⁶³ found that PROMs feedback increased the number of symptoms discussed during the consultation. However, Velikova *et al.*³²⁷ and Takeuchi *et al.*⁶³ found that the numbers of functional issues or non-specific symptoms discussed during the consultation were not different between the intervention and control groups. This suggests that the consultation was still focused predominantly on medical issues. Furthermore, in most cases patients were more likely to initiate the discussion of symptoms, suggesting that PROMs feedback did not change the traditional medical structure of the consultation.

Further evidence to support this theory also comes from a trial conducted by Berry *et al.*³⁵⁴ They found that clinicians continued to focus on medical issues regardless of whether these were actually a problem for the patient. They also found that some symptoms, such as those relating to sexual function, were discussed more frequently when reported at threshold. The authors suggest that PROMs may act as an icebreaker for these sensitive issues; however, such issues were still discussed relatively rarely compared with medical problems. Furthermore, clinicians also actively avoided talking about some issues, such as cognitive functioning, even when these were a problem for the patient.

Further theory testing: why does patient-reported outcome measures feedback not change clinicians' communication practices?

The trials reviewed in the previous section provide some hypotheses, or further candidate theories, that can be further explored in the synthesis to explain why PROMs feedback may not change clinicians' communication practices in the consultation.

Theory 18

A first hypothesis, suggested by Takeuchi *et al.*,⁶³ is that clinicians or patients may not consider it within their remit or may judge that it is not appropriate to discuss psychosocial or functional issues during the consultation.

Theory 19

A second hypothesis, suggested by Berry *et al.*,³⁵⁴ is that clinicians may prefer to focus on issues that they can do something about and avoid discussion of issues they feel unable to treat.

We now review a number of studies to test theories together. These include three main types of studies: (1) studies that have conducted surveys to explore patients' and clinicians' preferences and perceived responsibility for raising HRQoL issues during the consultation; (2) studies that have conducted interviews or focus groups with clinicians to explore their views about the use of PROMs in clinical practice and/or their perceived responsibility for addressing particular issues; and (3) studies that have provided a detailed qualitative analysis of oncology consultations between clinicians and patients.

Detmar *et al.*³⁵⁵

This study explored patients' and clinicians' preferences for raising and discussing a range of HRQoL issues during the consultation. Patients ($n = 382$) who had undergone two cycles of palliative chemotherapy for incurable cancer were sent a questionnaire exploring their preferences for discussing six areas of HRQoL: physical symptoms, general physical condition, daily activities, feelings, social activities and relationship with partner and family. Patients were offered three response categories: no, preferably not; yes, but only if my doctor raises the issue; and yes, I would like to discuss this issue. A total of 273 patients returned a questionnaire. Ten of the 12 medical oncologists in the cancer treatment centre participated in the study. They were sent a questionnaire that asked about their perceived responsibility for asking about four areas of patients' HRQoL: physical symptoms, general physical condition, daily activities and feelings. They were asked to respond using three response options: I consider this to be my primary task; it is partially my task and partially that of other health-care providers; and it is not my task. Clinicians were also asked who tended to initiate discussion about the six HRQoL issues included in the patients' questionnaires, using four response options: me, the patient, both of us equally or neither of us.

The authors found that patients' willingness to raise and discuss HRQoL issues during the consultation varied depending on the issue. The majority of patients (90.2%) wanted to discuss physical symptoms and conditions with clinicians. Two-thirds (67.4%) wanted to discuss emotional and daily activities but one-quarter (26.4%) would wait for the clinician to initiate discussion about these. Fewer patients wanted to discuss social functioning (44.3%) or their relationship with their family (52.7%) and would prefer to wait for the doctor to raise these issues. A significant minority did not wish to discuss social functioning (19.1%) or their relationship with their partner or family (18.7%).

With regard to doctors' preferences, the authors found that all doctors ($n = 10$) primarily saw it as their task to initiate discussions about the patients' physical symptoms and condition, but fewer saw it as primarily their task to initiate discussions about emotional functioning ($n = 6$) or daily activities ($n = 6$). In terms of doctors' self-reported behaviour, half ($n = 5$) reported that they largely initiated discussions about physical issues and half ($n = 5$) reported that they and the patient jointly initiated such discussions. However, none of the doctors reported that they initiated discussions about emotional functioning; six doctors reported that the patient initiated these discussions and four doctors reported that they and the patient shared the initiation of these discussions. Half of the doctors ($n = 5$) reported that neither they nor the patient initiated discussion about social functioning.

The findings of this study were based on a small sample of doctors, and examined their self-reported behaviour rather than their actual behaviour during consultations. Similarly, patients self-reported their preferences, which may be different from what actually happens during consultations. Furthermore, the wording of the questionnaire items to patients was ambiguous; indicating that one would like to

discuss an issue is not equivalent to indicating that one would *initiate* discussion of this issue, but the authors interpreted these data as if this was the case. However, the findings suggest that patients' and clinicians' perceptions about who is responsible for initiating discussion about HRQoL issues during the consultation vary depending on the issue. In terms of the theories under test, patients wanted to discuss and clinicians were willing to initiate discussions about physical symptoms. However, fewer doctors saw it as their primary task to initiate discussion of emotional issues and none reported initiating a discussion about these issues by themselves. The authors noted that for some psychosocial issues there may be a 'conspiracy of silence' because 'both doctors and patients are reluctant to raise issues without a clear signal from the other that it is appropriate or desired'. If issues are not discussed, doctors are less likely to take action to address them.

Taylor *et al.*³⁵⁶

This paper reports on an analysis of oncology consultations from an unpublished trial in which patients were randomised to either a control group, in which no PROMs completion took place, or an 'attention control' group who completed the EORTC QLQ-C30 but the results were not fed back to clinicians. The analytical focus of the paper was not on the findings of the trial itself; instead, data were treated as if they were from a cohort study. Of the 502 patients who were invited to participate, 356 consented to take part (71%), but, of these, 120 (34%) discontinued. Of the 236 patients who completed the study, 212 had four complete consultation recordings, and these formed the sample for the study. Thirty-six doctors agreed to take part in the study.

At baseline, patients completed the FACT-G questionnaire and a questionnaire to measure their preferences and perceptions of communication during consultations (PCC). This 14-item questionnaire was based on the questionnaire used by Detmar *et al.*³⁵⁵ to explore patients' preferences for discussing the six HRQoL issues described above. Taylor *et al.*³⁵⁶ added an item on memory and then seven items that asked patients about their perceptions of the extent to which these seven HRQoL items were discussed during the consultation. All consultations were audio recorded, but only the first consultations were coded to determine whether or not the issues in the EORTC QLQ-C30 were discussed, who initiated the discussion and if any actions were taken. Doctors also completed a version of the PCC on entering the study. This paper focused on reported patients' and doctors' preferences and perceived responsibility for initiating discussion, and actual discussion of emotional issues and social functioning.

With regard to emotional functioning, the majority of patients (59%) and doctors (75%) said that they would initiate discussions about emotional issues, 30% of patients would wait for the doctor to initiate discussion and 25% of clinicians would wait for the patient to initiate. Half of the doctors felt that they often or almost always discussed emotional issues, but only 18% of patients perceived that doctors had discussed them. In reality, emotional issues were mentioned in 58 out of the 212 consultations (27%) and led to a discussion in 24 (41%) of these instances. Emotional issues were more likely to be raised by the patient (49/58; 85% of instances). Doctors acted on the discussion in 13 out of 58 (22%) of these instances. There was no relationship between the severity of patients' problems and whether or not they were discussed.

With regard to social functioning, just under half of patients (44%) and around one-third of doctors (39%) said that they would initiate discussion of social functioning during the consultation. Around one-third of patients (39%) would wait for the doctor to initiate discussion and 11% would rather not discuss those issues. Just over half of doctors (58%) would wait for the patient to initiate discussion, while 14 (39%) indicated that they would initiate the discussion of social issues. Compared with doctors, patients perceived that social issues were discussed less frequently during the consultation.

Social issues were mentioned in 97 out of 212 consultations (more frequently than emotional issues) and this led to a discussion in 37 out of 97 (38%) of these instances. Social issues were more likely to be raised by the patient (68/97; 60%). Doctors acted on social issues in 4 out of 97 (4%) instances. There was no relationship between the severity of patients' problems and whether or not they were discussed.

In this study, a significant number of patients who were invited to participate did not take part or dropped out during the study, raising questions about the extent to which the sample of patients was representative of attendees at oncology outpatients. The authors observed that there were differences in the extent of disease between participants and non-participants, but no differences in other disease or demographic variables. Therefore, this study is more likely to represent the views of patients with less severe or less advanced disease. Audio recordings focused on only one visit, and communication about social and emotional issues may have occurred in other consultations and thus been missed by the analysis.

In terms of the theories under test, these findings suggest that clinicians and patients have different preferences for raising and discussing social and emotional issues. Both patients and doctors indicated that they saw it as more appropriate to raise emotional rather than social functioning issues. However, in reality, social functioning was more likely than emotional functioning to be discussed during the consultation. The analysis of the audio recordings of consultations indicated that, similar to the findings of Takeuchi *et al.*,⁶³ patients were more likely than clinicians to initiate discussions about emotional and social functioning. Even if these issues were mentioned, they led to a discussion in only 22% of instances for emotional and 38% of instances for social, and much less often resulted in clinicians taking action in response. Thus, patients' and clinicians' preferences and their self-reported communication behaviour differs from their actual communication behaviour during the consultation. This might be a function of a different understanding of social and emotional issues between the participants of this study and the authors. This suggests that it is important to assess communication behaviour through the analysis of recordings of consultations, rather than through self-reported behaviour.

The authors offered a number of reasons that emotional issues were not discussed. First, medical issues may take precedence in consultations that focus on reviewing treatment effectiveness, which is still largely determined in terms of tumour size. Here, emotional issues may seem less relevant to doctors' decision-making. Second, patients may not feel that it is the doctors' role to address emotional issues. Third, doctors may not discuss emotional issues because they feel that they cannot resolve them. Doctors may be more comfortable discussing social issues because these are more likely to be linked to medical issues, which doctors feel more comfortable dealing with.

Velikova *et al.*³⁵⁷

This study examined views of the use of PROM questionnaires among cancer patients and clinicians at the Leeds Cancer Centre. Focus groups were held with 31 patients, selected by age (< 50 years and ≥ 50 years), disease stage (early disease, treated with curative intent, or advanced disease, on palliative treatment) and hospital site (eight groups), plus one group involved in an earlier quality-of-life questionnaire study. Focus groups were also held with 16 oncologists, divided into four groups according to specialty (medical oncology and clinical/radiation oncology) and qualification level (specialist registrars and consultants). In addition, 20 patients who were unable to attend the focus groups completed a postal questionnaire. Participants were shown an example quality-of-life questionnaire (EORTC QLQC-C30), graphs showing scores on pain and fatigue and photographs of the computer touchscreens used to administer the questionnaire in clinics.

The findings cover a range of issues affecting the perceived value of PROMs, such as content, timing and presentation. We focus here on issues raised in relation to theory 18, on clinician remit, and theory 19, that clinicians avoid discussion of issues considered untreatable. However, we start by noting that patients and clinicians also expressed views and ideas that support the theories reviewed in *Chapter 8*, that being asked to complete a PROM signalled that 'people were still interested in me . . . people still wanted to know'.³⁵⁷ Furthermore, patients commented that their willingness to share information depending on how easy they found the consultant to talk to and noted that 'there are consultants around who are not so easy to talk with . . . it's hugely dependent on the consultant and the patient'.³⁵⁷ Consultants also expressed some concern that PROMs may damage the clinician–patient relationship because 'for the patient it feels like you are taking away the opportunity to talk to the doctor properly'.³⁵⁷

With regard to the theories under review in this chapter, the findings suggest that emotional issues are not routinely discussed in consultations with doctors. As explained by one oncologist:

The things we do tend to miss out are things like . . . stress and anxiety, depression. Unless they actually look depressed or they volunteer something, we don't [discuss].

*Physician consultant*³⁵⁷

Patients also reported that psychological effects were not discussed:

I did crumble emotionally, and that wasn't addressed.

*Patient young early disease*³⁵⁷

The findings provide some support for the role of remit and whether or not issues are seen as treatable as contributing to this limited discussion of emotional issues. In relation to theory 18, on clinician remit, the findings suggest that patients may see specific issues as appropriate for discussion with specific clinicians. The authors report that 'Some patients felt they only want the oncologists to treat the cancer and preferred to discuss emotional, family or social issues with nurses'.³⁵⁷ For example:

I see my oncologist as sort of an expert in cancer treatment and so the questions focus on cancer.

*Patient young early disease*³⁵⁷

Who should ask about personal issues, should the oncologist do it? How many hats does an oncologist have to wear?

*Patient young early disease*³⁵⁷

The authors reported that both patients and clinicians saw quality-of-life issues as better suited to discussion with nurses: 'Patients and professionals suggested that the information on practical problems such as social benefits and finances is more appropriate to be fed back to nurses'.³⁵⁷

In relation to theory 19, clinicians expressed concerns about PROMs raising topics that they could not treat:

Do we want to bring up a lot of things that we can't do anything about?

*SpR (specialist registrar)*³⁵⁷

In assessing this evidence on theory 18, the idea that consultations with oncologists should focus on more medical aspects, while broader quality-of-life issues would be more suitable for discussion with nurses, appears from the article wording to have been shared by at least some patients and clinicians. However, direct evidence, provided in the form of quotations, came from patients rather than clinicians. In addition, the findings focused on consultations with specialised hospital doctors, and it may be that the idea of leaving quality-of-life discussion to nurses does not apply in other settings (e.g. with GPs). Nevertheless, the findings provide support for the idea that whether or not PROMs issues are discussed will depend partly on whether clinicians see these issues as part of their remit and patients see these issues as suitable for discussion with a specific clinician. This article provides only limited evidence on theory 19, with just one quotation and no further discussion specifically about patient concerns being considered untreatable.

Absolom et al.³⁵⁸

This study explored the perceptions of cancer doctors and nurses about the detection and management of emotional distress. Interviews were conducted with six clinical nurse specialists, eight oncologists, four surgeons and five ward sisters from hospitals in Yorkshire, including a cancer centre and smaller district hospitals. Framework analysis was used to interpret the data. Clinicians were asked about their current roles and responsibilities in the detection and management of emotional distress, the use of screening tools and providing access to expert psychological support. The article provides information on the perceived value of screening tools. However, we focus here on the findings in relation to which kind of professionals are considered responsible for addressing emotional issues to test theory 18 on clinicians' perceptions of their remit.

The authors report that 'All staff accepted responsibility for the detection of ED [emotional distress]'.³⁵⁸ However, the detection and management of emotional distress was more often undertaken by nurses, and was seen more clearly as part of their remit. The authors note that 'Almost all oncologists and surgeons reported a reliance on the CNS [clinical nurse specialist] or breast care nurses in detecting distress',³⁵⁸ and that 'Oncologists and surgeons gave priority to cancer treatment and generally did not consider the management of ED to be a key part of their role'.³⁵⁸ Emotional distress was seen as outside doctors' remit: 'generally oncologists and surgeons do not see detection as part of their day-to-day role and do not routinely explore unless it influences treatment plans or if the patient explicitly discloses problems'.³⁵⁸ For example, one oncologist is quoted as saying:

I wouldn't go digging any deeper without some push to do so.

011, consultant oncologist, breast DGH (district general hospital)/CC³⁵⁸

Another suggested that:

I really don't think it's a surgeon's remit to start probing other issues, other paths, when in fact their main concern is that they have a cancer and they're coming in for an operation.

014, consultant surgeon, breast/thyroid, DGH (district general hospital)³⁵⁸

The authors suggested that doctors felt that focusing on emotional issues could 'adversely impact the medical management of cancer and be time-consuming'.³⁵⁸ In contrast, the authors observed that, 'Nursing staff appeared comfortable exploring emotional issues'.³⁵⁸ Comments from nurses indicated that they are left with responsibility for identifying and discussing emotional issues:

[F]rom what I've seen the medical staff tend to shy away from sitting and talking and exploring reasons why people are emotionally distressed, the emotional problems, that's very much left to the nursing staff to deal with.

021, ward sister, surgical, lung, CC³⁵⁸

Remit appeared to vary with the nature of emotional distress. The authors report that 'Oncologists and surgeons felt able to deal with the distress related to clinical problems and uncertainties',³⁵⁸ for example by providing more information about treatment options. The reliance on nurses seemed to apply more to wider emotional issues.

The article also indicated aspects of context that affected clinicians' abilities to discuss and manage emotional distress, including time and access to services for referral. On the former, 'all the professional groups viewed time pressures to be a fundamental problem impeding the management of distressed patients'.³⁵⁸ This was affected partly by the availability of clinical nurse specialists and variations in workload. On service availability, the authors reported that 'One of the main issues staff perceived as a barrier to the successful management of ED was poor access to specialist psychological and supportive services'. Access to services varied between hospitals, but, in general, psychology services were seen as overburdened, which made some clinicians reluctant to refer patients:

[T]he last time I tried to refer someone I was just told there was such a long waiting list I was told they couldn't be seen. I don't think I've explored it again.

022, surgical ward sister, DGH (district general hospital)³⁵⁸

As well as a shortage of services, patient referral was hindered by clinicians' lack of knowledge about appropriate referral pathways, particularly for less severe cases.

This study relied on clinicians' reports of who discussed and managed emotional distress. However, the sample included a range of professional roles, and comments suggested shared perceptions among participants, strengthening the evidence. The findings support suggestions from other studies that

emotional issues are seen as outside the remit of some clinicians, and in particular that nurses may be more likely than doctors to discuss and address emotional issues. The findings also point to the importance of time, access to services for referral and knowledge of referral options as influences on the discussion and management of emotional distress.

Annells and Koch²⁴⁴

The Annells and Koch²⁴⁴ study, described in relation to patient completion of an individualised PROM in *Chapter 8*, provides an example in which clinicians do discuss issues covered by the PROM with patients, including those that they see as outside their usual remit and untreatable. As explained in *Chapter 8*, the study reported the experiences of eight district palliative care nurses during a pilot study to explore the feasibility and acceptability of using an individualised PROM, the CGI, during the first assessment of patients admitted to the palliative care service.

The CGI was administered by nurses through an interview with patients. Clients' concerns, as raised through the PROM, were often seen as beyond the scope of nurses' standard remit or as problems that they could not treat, for example more emotional or social issues. The authors report that the nurses 'were disconcerted by these issues being raised as they felt that classically they were not issues that district nurses considered to be within the scope of nursing, that is, not relating to nursing intervention that is the concrete, physical doing of something'.²⁴⁴ For example, one nurse is quoted as saying that:

*They were things that traditionally nurses don't focus on. Well, we feel that we can't do anything about it, so we don't focus on them, I think. We develop in a system that is so task and time conscious that if you are not 'doing' something, then it has no value.*²⁴⁴

Another quotation from a nurse makes the same point about focusing on issues that can be solved:

*It isn't fixing the problem like we do . . . Nurses are reluctant to establish care plans for issues about which it is thought they can do little or nothing to resolve.*²⁴⁴

These findings suggest that nurses feel that their role is to 'solve' problems for patients. Findings from the PROM discussions did, in some cases, lead to interventions, including actions on physical symptoms such as bowel management but also action in response to wider quality-of-life and emotional or social issues, for example referral for depression, consideration of quality of life in treatment options and practical support such as assistance with funeral planning. However, sometimes the patients' concerns were simply discussed; the nursing advisors felt that this was often valuable in itself, even if further action was not possible. As discussed in *Chapter 8*, patients appeared to experience emotional benefits from discussing their concerns and being listened to. One nurse suggested that:

*It's not easy to accept that some problems you can't fix and you just listen to what the person is going to say and by understanding . . . what they are going through, that is in itself useful and it helps that person to come to some realisation and acceptance of what's going on.*²⁴⁴

Although this support for more emotional or social issues and listening to problems that could not be resolved was presented by nurses as outside what is typically seen as standard care work, the authors reported that the nurses also understood these aspects of relationship building and listening as an important part of what nursing actually involves: quoting one participant, the authors noted that 'despite these aspects not being commonly recognized nor documented, "It is nursing. We all know what it is"'.²⁴⁴ Furthermore, one nurse explained how the use of an individualised PROM had taken her away from the style of questioning she usually used:

*You cease to be a nurse a little bit . . . a nurse is like 'How is your pain? Are you nauseous? Are you constipated?' You know, bang, bang, bang then out the door!*²⁴⁴

This was a small pilot study. However, the 40 weeks of PROMs use provide a relatively long period for the nurses to gain experience of the tools, and the findings provide further evidence to test ideas about clinicians' use of PROMs data. In terms of the theories under test, this study suggests that, in some situations, clinicians are willing to discuss issues raised through PROMs that they see as beyond their standard remit and untreatable. In this study, the nurses felt that just discussing issues had a value. There are several aspects of the context that might have encouraged the discussion of sensitive issues that could not necessarily be solved or that might have been seen as outside clinicians' remit. One aspect specifically addressed in the article is the role of nurses (rather than doctors). Although the nurses felt that listening to patients and relationship building was not typically seen as part of their work and was not formally documented in care plans, they also emphasised that this kind of support is part of what nurses do. This idea that nurses may be more likely to spend time on discussion of emotional and social issues or problems they cannot solve fits suggestions from studies by Velikova *et al.*³⁵⁷ and Absolom³⁵⁸ that patients and clinicians sometimes see broader quality-of-life issues as part of nurses' remit, rather than work for doctors. A second aspect of context is the focus on palliative care, where providing emotional support might be considered a more important part of clinicians' work, with less attention given to active treatment and cure. A third important aspect of this context is that there may have been time available for discussion. Although this is not specifically addressed in the article, there may have been more time for discussion in the context of district nursing and initial assessment visits, than, for example, during short outpatient appointments. Finally, as reviewed in detail in *Chapter 8*, the use of an individualised rather than a standardised PROM may have served to broaden the scope of the discussion.

Greenhalgh *et al.*⁶⁵

This paper reports on a secondary analysis of data collected as part of the three-arm RCT conducted by Velikova *et al.*³²⁷ The analysis focused on understanding the dynamics of the clinician–patient interaction in more detail to explore *how* clinicians referred to HRQoL data during the consultation. The specific aims were to examine how these data were referred to in the process of eliciting patient's concerns, making decisions about supportive treatment and making decisions about systemic treatment. Conversation analysis was used to examine 18 consultations from the intervention arm and four consultations from the attention control arm. The consultations used in the analysis were purposively selected from the intervention arm of the trial using sequential sampling criteria. These were consultations in which doctors received each patient's PROMs data for the first time, and consultations in which the impact of chemotherapy was reviewed or discussed and where PROMs data were explicitly referred to. Consultations were selected from this set to reflect interactions with patients who had varied PROMs profiles. In addition, consultations from the attention control arm in which patients had high fatigue were selected because discussion of fatigue was identified as a theme during the analysis.

In consultations for which PROMs data are available, the clinician is in a position of having to manage and perhaps adjudicate between two sources of information – patients' verbal reports and PROMs data – which both claim to represent the patient's subjective perceptions of the impact of their disease and its treatment. The extracts from consultations presented in the paper suggest that clinicians were engaging in a process of reconciling differences between PROMs data and patients' verbal reports and checking the validity of PROMs data. For example, one patient indicated in their verbal report that apart from a bout of tonsillitis he had been 'ok'; however, the PROMs data showed a number of high scores, suggesting that he was experiencing a number of problems. The clinician directly pointed out this difference to the patient and in turn the patient accounted for why the scores were high, which revealed a number of causes that were not related to the cancer or chemotherapy. The authors commented that certain PROMs scores, particularly emotional functioning, did not distinguish between 'problems related to cancer or its treatment' and 'other problems', and thus exploration is required to account for the cause of the high scores.

In another example, the clinician mentioned that the patient had scored highly on a number of PROMs and ran through each one, inviting the patient to account for them. These accounts revealed that many issues were either not related to the cancer treatment or not problematic for the patient and, as such, were not pursued further by the clinician or patient. The authors observed that high PROMs scores did not

necessarily lead to discussion or action if a patient's account revealed that these were not related to the cancer or its treatment or were not a problem for the patient. In terms of the theories under test, these findings indicate that, in exploring patients' problems, clinicians did not take the PROMs scores at 'face value' but attempted to verify them against patients' verbal reports. When these reports revealed that the issue was not related to cancer or its treatment or was not a problem for the patient, discussion was closed down, suggesting that clinicians see their remit as focusing on issues that are related to cancer or its treatment.

However, even when issues were problematic for patients, clinicians did not always pursue discussion or take action to address them. The authors found that patients were more likely to be offered symptomatic treatment for sickness, constipation, pain and depression (for which effective treatment was available), but much less so for fatigue (for which there is often no treatment). Discussion was constrained by the doctor offering a 'candidate cause' for the problem, such as the chemotherapy, which removed the need for the patient to account for it, but, in doing so, limited exploration of its effect on the patient. They were also closed down by changing the subject or presenting fatigue as something to put up with. The authors hypothesise that discussions about fatigue were closed down because clinicians did not know how to or felt that they could not offer a treatment or solution to the problem.

Finally, the authors found that treatment and chemotherapy regimens were changed if treatment was not reducing tumour size. In instances when the treatment was working, chemotherapy regimens were only changed to address a collection of severe side effects, provided that changes to the dosage or frequency of the regimen were not likely to reduce its effectiveness. PROMs data were more likely to be used to support rather than determine these decisions and, in these circumstances, clinicians often referred to PROMs data as if they were another test result. Thus, PROMs data were used to legitimise treatment decisions rather than determine them.

This was a small study in which only 22 out of the 792 consultations in the trial were analysed. Although the authors indicate confidence that data saturation was reached, it is possible that different patterns would have emerged had a larger number of consultations been subjected to detailed analysis. The authors also acknowledge that caution is needed in generalising their findings. In relation to the theories under test, the findings of this study highlight that the discussion of PROMs data during the consultation creates a number of interactional problems for clinicians. The findings also indicate that PROMs data do not speak for themselves, and clinicians attempted to reconcile differences between PROMs data and patients' verbal reports and explored the cause of high PROMs scores by inviting patients to account for their scores. These accounts sometimes revealed that patients' problems were not related to cancer or its treatment and discussions were closed down. This may explain why Takeuchi *et al.*⁶³ found that discussion was related to the severity of symptoms but not functional problems, as functional problems may be more likely to reflect problems not related to cancer or its treatment. This further suggests that clinicians view their remit as dealing with issues related to cancer and its treatment, in line with findings from Velikova *et al.*³⁵⁷ and Absalom *et al.*³⁵⁸

The findings also suggest that discussion is more likely to occur when it is about symptoms that doctors feel they can treat. In line with the findings by Berry *et al.*,³⁵⁴ clinicians avoided or curtailed discussion of issues for which they saw little scope for action, in turn reducing changes to treatment or other kinds of management responses that could address issues highlighted through the PROM. This may be a more significant issue in contexts in which clinicians lack the understanding of management options for issues, such as fatigue, that are seen as harder to address. The findings also show that action in response to PROMs data will depend on the fit between PROMs feedback and other information and treatment priorities. Decisions about patient management consider a number of sources of information, such as findings from clinical investigations and clinical judgement, and clinical priorities (such as tumour progress) may dominate decision-making.

Rogers *et al.*³⁵⁹

This study aimed to explore how patients and clinicians discuss patients' problems and concerns during outpatient oncology consultations. The study does not specifically focus on the use of PROMs data in the consultation, but does provide useful evidence to test theories 18 and 19. Data were collected during outpatient clinics of a regional teaching hospital. Seventy-four consultations between 39 cancer patients and 15 doctors were observed, audiotaped and analysed using a combination of qualitative content analysis and conversation analysis. In addition, field notes were also taken during the observations. The analysis focused on understanding how pain was discussed between patients and clinicians.

The authors found that pain was identified as a substantial topic of conversation in 39 of the 74 consultations. The authors noted that pain (like fatigue) is a non-specific symptom and clinicians in their study focused on determining whether or not the pain experienced by patients was related to cancer in order to avoid becoming engaged in discussion about pain or problems perceived to fall outside their remit. The authors observed that clinicians controlled the consultation agenda and their questioning about pain was guided by 'a heuristic' to determine if patients were reporting 'the right kind of pain' (i.e. pain that was specifically related to cancer). For example, patients with pelvic tumours were specifically asked about abdominal pain, swelling and bowel function, whereas patients with lymphoma were questioned about 'lumps and bumps', fever, sweating and weight loss. If the patients' reports of pain were suspected as being related to cancer, the patient was questioned further. However, when patients' reports of pain suggested that this was not related to cancer, clinicians closed down the discussion using a number of interactional strategies, such as interrupting the patient, minimising symptoms, changing the topic and privileging objective evidence over patient reports. In their discussion, the authors noted that use of these information limiting strategies 'helped to maintain not only the focus on the physical domain but on pain that was amenable to a "technical fix"'.³⁵⁹ They also argued that their findings showed that clinicians 'defined their roles narrowly as providing specialist cancer services only and not symptom management per se'.³⁵⁹

This was another small qualitative study in a single centre. Little information is provided about the reasons for the observed consultations, for example whether they were initial referrals, patients on active treatment or follow-up consultations. This is important missing contextual information, as clinicians' interactional strategies may vary depending on the purpose of the consultation. Nonetheless, the study provides useful evidence to understand how clinicians' perceptions of their remit can influence the interactional strategies used during the consultation to close down and open up topics for discussion. It suggests that clinicians avoid becoming involved in discussing symptoms that are perceived as not being related to cancer or its treatment. It also suggests that they perceive their remit as detecting and managing the patient's cancer, rather than symptom management per se. This may be different from consultations with patients undergoing active treatment for cancer, during which greater attention may be paid to symptom management.

Detmar *et al.*³⁶⁰

This study investigated the primary reasons oncologists discontinued or modified treatment for patients receiving palliative chemotherapy and explored whether or not these decisions were related to clinicians' perceptions of patients' HRQoL. This study does not examine the role of PROMs feedback in these decisions, but does provide evidence to explore the extent to which clinicians' decisions about chemotherapy are influenced by biomedical or psychosocial issues.

The study included 203 patients receiving outpatient palliative chemotherapy. Four consecutive consultations were audio recorded and analysed using a coding checklist to identify whether treatment was stopped or modified (e.g. dose reduced, or treatment postponed or changed to a milder regimen). When treatment was stopped or modified, the primary reasons were coded as tumour progression, treatment toxicity, HRQoL considerations or combination of these. Toxicity was defined as on the basis of evidence from physiological tests, for example blood values or liver function, while HRQoL issues were defined as symptoms such as fatigue, pain or nausea. A total of 704 tape-recorded consultations were included in the analysis. Clinicians' perceptions of the patients' HRQoL were assessed after each consultation using the COOP/WONCA health assessment charts.

The authors found that treatment was given as planned in 54% of cases, modified in 26% and discontinued in 20%. Based on the consultation discussions, the main reason for changing treatment was toxicity (22 cases), followed by HRQoL issues (18 cases) and tumour progression (14 cases). The main reason given for discontinuing treatment was tumour progression (23 cases), followed by HRQoL issues (six cases) and treatment toxicity (three cases). Treatment decisions varied mainly with tumour status and treatment toxicity.

The findings also revealed how the relative importance of HRQoL varied between patient contexts. Treatment was always modified or discontinued for patients with tumour progression or serious toxicity, regardless of HRQoL; the authors state that 'in the presence of tumour progression or serious toxicity, HRQoL considerations played little or no role in treatment decisions'.³⁶⁰ When there was no evidence of tumour progression or serious toxicity, HRQoL played a stronger role: in these situations, treatment decisions varied significantly with HRQoL ratings. However, even here, HRQoL concerns influenced only a minority of cases: 'approximately 70% of patients without evidence of tumour progression or toxicity, but with seriously impaired HRQoL, continued to receive their treatment as planned'.³⁶⁰ In addition, HRQoL was a more frequent influence on decisions to modify treatment for patients receiving second-line (or more) chemotherapy than for those receiving first-line chemotherapy, although this was not statistically significant.

In conclusion, the authors stated that 'HRQL considerations, in and of themselves, seem to trigger changes in treatment in only a minority of cases'.³⁶⁰ The authors suggested possible explanations for the limited role of HRQoL issues in treatment decisions. These included a dominant medical ethos against stopping active treatment because clinicians are 'trained to intervene'.³⁶⁰ In addition, clinicians might have felt that stopping chemotherapy would bring further deterioration in HRQoL. A further possibility is that patients may accept HRQoL limitations for additional survival time.

One limitation of this study is that treatment decisions may take quality of life issues into account without this being explicitly indicated in consultation discussions. However, if clinicians are giving equal weight to HRQoL issues in their decisions, it is not clear that they would be less likely to verbalise these rationales than those related to issues such as toxicity. In terms of the theories under test, these findings support the theory that clinicians may focus management decisions on biomedical considerations of tumour progression and treatment toxicity and that HRQoL issues play a relatively minor role in treatment decisions. This is in line with findings from the Greenhalgh *et al.*⁶⁵ study that changes to patient management tended to reflect issues such as toxicity rather than the broader concerns raised through PROMs. Detmar *et al.*'s³⁶⁰ findings suggest that quality of life concerns can have more influence on treatment when there are no other over-riding issues. Although the article does not focus on the quality of discussion or use of PROMs, it provides evidence that can help to explain why PROMs, particularly those focused on quality of life issues, may not lead to changes in patient management.

Chapter summary

The studies reviewed here used a range of study designs to test our theories. Through these additional theories we sought to understand and explain why PROMs feedback in oncology settings may not change clinicians' communication practices. We summarise the findings of these studies in relation to our theories.

Theory 18

We found evidence to support but also refine this theory. Detmar *et al.*³⁵⁵ found that both patients and doctors saw physical symptoms as clearly within the doctor's remit and would be willing to discuss them during the consultation, whereas none of the doctors would be willing to initiate discussion about emotional issues on their own. Similarly, Taylor *et al.*³⁵⁶ also found that although a substantial percentage of patients and doctors felt it appropriate to raise discussion about emotional issues, in reality such issues were mentioned in only 27% of consultations and led to a discussion in fewer than half of these instances. Although fewer patients and clinicians felt that it was appropriate to raise social functioning issues during the consultation, these issues were actually raised more frequently (46%) than emotional problems.

The authors hypothesised that this was because problems with social functioning were more likely to be caused by the physical impact of cancer, which oncologists saw as the focus of both their role and the consultation.

Interview and focus group studies with clinicians and patients^{357,358} also suggest that patients see the doctor's remit as focusing on physical issues and are unsure if it is the doctor's role to address psychosocial issues. Doctors also acknowledged that emotional issues are not routinely discussed during consultations and that they would not enquire about them unless patients volunteered such information. Doctors, especially surgeons, felt that their remit was treating the patient's cancer and, although they felt able to deal with emotional issues related to clinical problems, they felt that it was the nurses' role to deal with wider emotional issues. Nurses also perceived that much of the responsibility for dealing with emotional issues also fell to them. These findings suggest a subtle revision to our theory: doctors do feel that certain issues, such as emotional issues, fall outside their remit, but they also perceive that it is their job to deal with issues that are specifically related to the patient's cancer and its treatment, rather than wider problems in the patient's life. Thus, physical symptoms can be more clearly attributed to cancer or its treatment, but emotional or functional issues may have a variety of causes, many of which doctors feel it is not within their remit to address.

We also found some evidence to show how doctors' focus on dealing with issues directly related to cancer and its treatment influenced their interactions with patients and also their treatment decision-making. Rogers *et al.*³⁵⁹ found that oncologists in outpatient clinics used a line of questioning during their consultations that sought to determine whether or not patients' reports of pain were specifically related to cancer and its treatment. In situations when they suspected that the pain was not related to cancer, they closed down the discussion using a number of interactional strategies, such as interrupting the patient, minimising symptoms, changing the topic and privileging objective evidence over patient reports. Greenhalgh *et al.*⁶⁵ observed how doctors' focus on identifying problems related to cancer and its treatment also influenced how doctors referred to PROMs data during their consultations. The authors noted that PROM scores do not distinguish between 'problems related to cancer' and 'other problems', and so doctors engaged in a process of determining whether or not high PROMs scores reflected issues that were specifically related to cancer. This involved reconciling differences between PROMs scores and patients' verbal reports and inviting patients to account for high PROMs scores. Although these strategies opened up discussions between the patient and the doctor, these discussions tended to be closed down when patients' accounts revealed that the issue was not problematic for them or was not related to cancer. This suggests that doctors implicitly see PROMs scores as relevant in so far as they provide information that reflects the impact of cancer and its treatment.

Detmar *et al.*³⁶⁰ also showed that doctors' decisions to modify or stop chemotherapy for patients with incurable cancer were more influenced by tumour progression and toxicity than by HRQoL issues. Similarly, Greenhalgh *et al.*⁶⁵ also found that decisions to stop or modify systemic treatment were largely based on tumour progression. Treatment regimens were only changed in response to a cluster of debilitating side effects and only if such changes would not affect the likely impact of treatment on the tumour. This suggests that, when it comes to making decisions about chemotherapy, doctors rely much more on evidence that more clearly represents the impact of treatment on the biomedical aspect of cancer, rather than on HRQoL information that may reflect not just the impact of cancer and its treatment but other aspects of patients' lives. It also indicates that doctors focus on prolonging life, rather than maximising quality of life, even in situations when cancer is incurable.

Theory 19

We found some evidence to support but also to refine this theory. Greenhalgh *et al.*⁶⁵ found that when patients' verbal reports confirmed that high PROMs scores were likely to be due to cancer or its treatment and were problematic for the patients, patients were much more likely to receive supportive treatment for symptoms such as nausea or constipation than for non-specific issues such as fatigue. This was not simply because doctors perceived that the fatigue was not related to cancer or its treatment; in a number of

instances the doctor offered chemotherapy as a candidate cause for the fatigue, thus clearly demonstrating its link to cancer treatment, although, at the same time, constraining patients' accounts of this symptom. Rather, Greenhalgh *et al.*⁶⁵ hypothesise that this was because there are few physical treatments doctors can offer to address this symptom. As such, discussion of fatigue was more likely to be shut down because this was presented as an inevitable side effect of chemotherapy that patients had to tolerate to benefit from either the curative or the life-prolonging benefits of systemic therapy. This suggests that doctors preferred to focus discussion on issues they felt that they could treat. Rogers *et al.*³⁵⁹ also argued that limiting discussion of pain to that directly related to cancer served to focus the consultation on pain that was amenable to a 'technical fix'. Similarly, Velikova *et al.*³⁵⁷ found that doctors were concerned that PROMs may raise issues that they were powerless to change.

This need to 'do something' may be a function of the ways in which clinicians are trained and how they interpret their professional role. The nurses in Annells and Koch's²⁴⁴ study commented that their role was to 'fix things' and that they worked within a system, where, unless they were 'doing something', it had little value. They reported that nurses were unlikely to develop care plans for issues that they felt they could do little to resolve. Consequently, nurses reported feeling uncomfortable when the use of an individualised PROM, the CGI, raised issues that they felt were traditionally outside their remit. However, through the continued use of an individualised PROM, the nurses reflected that, although it was not easy to accept there were some problems they could not fix, they had come to realise that sometimes just listening to the patient discuss these issues could have therapeutic benefits. The nurses further reflected that this was an important part of nursing practice. Nurses' willingness to discuss issues they felt unable to treat may be a function of a whole configuration of contextual factors. First, as previously discussed, both nurses themselves and other professionals perceived that it was the nurse's role to address the wider aspects of patients' lives.³⁵⁸ Second, this study took place in palliative care, where the focus is on maximising patients' quality of life, rather than their quantity of life. This is different from outpatient oncology consultations, where many patients are undergoing active treatment with curative intent. Thus, the focus of consultations in that context is on determining whether or not the treatment is achieving these goals and thus largely relies on information from biomedical tests. Finally, in this study,²⁴⁴ an individualised PROM was used. Our synthesis in *Chapter 8* suggested that individualised measures were more useful than standardised PROMs as a 'conversation opener' as they enabled patients to tell their story and thus may have been more likely to have raised and perhaps facilitate the discussion of issues that were not necessarily amenable to treatment.

Thus, these findings suggest a refinement or reframing of our original programme theory. Our original programme theory was that PROMs feedback would result in increased discussion of HRQoL issues during the consultation, and would promote increased patient well-being or improve patients' experience of the consultation. However, our synthesis suggests that it is assumed that PROMs feedback to clinicians will shift the focus of the consultation away from traditional biomedical issues towards more psychosocial issues. This is particularly the case in oncology settings but may also be the case for other long-term conditions. However, our synthesis suggests that underpinning this programme theory is an assumption that PROMs feedback to clinicians will trigger a shift in clinicians' perceptions of their professional remit. Thus, in oncology, doctors will shift their focus on defining treatment success in biomedical terms towards considering and addressing the impact of treatment on psychosocial issues also. However, the evidence reviewed in this chapter suggests that although PROMs increases the discussion of symptoms that are clearly related to cancer and its treatment, patients' problems that are perceived to be outside the clinician's remit, or which they feel they cannot treat, receive less discussion and are less likely to be addressed. Instead, doctors engage in a process of determining whether PROMs scores are related to cancer (and thus within their remit to address) and close down discussion of issues that are not related to cancer or where they cannot intervene. Professionals' perceptions of the remit and their role are developed over years of training and are rooted in and mutually reinforced by the complex division of labour across different professional groups. These social structures are likely to require more than just the feedback of PROMs scores in order to change them.

Chapter 10 Discussion and conclusions

In this chapter, we summarise and discuss the findings to emerge from our review and provide recommendations for research and practice. We start by summarising the key findings, then consider the strengths and limitations, and discuss implications for research and practice based on our conclusions.

Summary of main findings

We have already provided a detailed summary of these findings at the end of each chapter, so here we focus on highlighting key messages in the report.

We start by providing a recap of the aims and objectives of our review:

1. To identify and classify the various ambitions for the use of PROMs data at the aggregate and individual level:
 - i. produce a comprehensive taxonomy of the 'programme theories' underlying these different functions, and capture their differences and the tensions that may lie between them
 - ii. produce a logic model of the organisational logistics, social processes and decision-making sequences that underlie the collation, interpretation and utilisation of PROMs data.

We addressed these aims and objectives in *Chapter 3* for aggregate PROMs data and *Chapter 7* for individual-level PROMs data.

2. To test and refine these programme theories about how PROMs feedback is supposed to work against existing evidence of how they work in practice to:
 - i. identify the implementation processes that support or constrain the successful collation, interpretation and utilisation of PROMs data
 - ii. identify the mechanisms and circumstances through which the unintended consequences of PROMs data arise and those where they can be avoided.

We addressed these aims and objectives in *Chapters 4* and *5* for PROMs use at the aggregate level and in *Chapters 8* and *9* for PROMs use at the individual level.

We now summarise the key issues to emerge from the theory elicitation and synthesis findings.

Chapter 3: a taxonomy of the programme theories underlying the use of aggregate patient-reported outcome measures data

The feedback of aggregate PROMs data shares many of the underlying ideas and assumptions found in feedback and public reporting of other forms of performance data, such as mortality report cards, hospital star ratings and patient experience data. We also observed that feedback of all performance data has been implemented against a background of many other initiatives designed to improve the quality of patient care, such as the NHS Outcomes Framework and financial incentives such as the QOF in primary care and CQUIN and BPTs in secondary care. They have also been introduced into health systems that have seen a shift towards greater patient involvement and choice, and the provision of information to patients in order to support these choices. These policies and programmes have influenced the development and aspirations of the feedback of aggregate PROMs, and shaped the ways in which this feedback leads (or does not lead) to its intended outcomes.

We then provided a comprehensive taxonomy of the intended functions of PROMs feedback, which include:

- support for patient choice
- accountability to stakeholders
- comparison of provider performance
- assessment of the appropriateness of surgery.

These aims are expected to stimulate providers to improve patient care through a range of different mechanisms, including:

- having an intrinsic desire to improve
- protecting their professional reputation
- protecting their market share
- being as good as or better than their peers.

We also identified some potential unintended consequences of PROMs and performance feedback such as gaming, tunnel vision, effort substitution or avoiding high-risk (sicker) patients. We also described a number of potential contextual configurations that may influence the mechanisms through which PROMs feedback works and its outcomes. These included:

- the credibility of data
- the extent to which PROMs and performance data are 'actionable'.

This analysis informed the programme theories tested in *Chapters 4 and 5*.

Chapter 4: the mechanisms through which the feedback of patient-reported outcome measures and performance data works

We tested a number of theories about the mechanisms through which providers respond to PROMs and performance data:

- *Theory 1* The feedback of PROMs or performance data leads to improved patient care.
- *Theory 2* Feedback influences provider behaviour according to whether it is delivered privately or publicly.
- *Theory 3* Public reporting of poor performance threatens providers' market share and provokes a significant provider response.
- *Theory 4* Providers perceive that report cards damage their reputation.
- *Theory 5* Providers wish to be as good as or better than their peers.
- *Theory 6* Patients choose hospitals on the basis of quality information.

Owing to the paucity of evidence evaluating the use of PROMs as a tool to improve patient care, much of the evidence to test these theories was drawn from the public reporting and feedback of other performance data. There is evidence to indicate that public reporting places additional pressure on providers to respond, particularly poor performers. In general, hospitals do not see a change in their market share following the public reporting of hospital quality. However, they are more likely to experience a change in market share if information about service quality reveals that previously 'good' hospitals have performed poorly (i.e. if performance data damages their reputation). Providers perceive that the public reporting of poor performance damages their reputation and so they take action to improve patient care in response to this. They perceive that patients' choice of hospital is influenced by the media reporting of poor performance, whereas patients are more influenced by media reporting of 'untoward' adverse events. Providers are also motivated to respond to public reporting in order to be as good as or better than their peers, which in turn improves their reputation and, potentially, their market share. Patients do not use publicly reported information about service quality to inform their choice of hospital, but instead rely on their personal experience, the opinions of friends and family and advice from their GP.

The different reasons providers might respond to performance data are not mutually exclusive but rather are inter-related. Providers take steps to improve the quality of patient care in order to provide patients with a good experience of care and to maintain their loyalty. As a result, patients will tell their family and friends, thus upholding the hospital's good reputation and making it more likely that they too will choose to go to the hospital. Providers also take steps to be as good as or better than their peers, which in turn enhances their reputation with referring clinicians, who are then more likely to refer patients to them, which in turn protects their market share. Publicly reported quality information itself plays little or no role in patients' choices of hospital or in GPs' decisions to refer patients. Performance data are perceived as a threat to a provider's reputation if the provider and their service quality are misrepresented or misreported in the media, which could, in the eyes of providers, damage their reputation.

Chapter 5: contextual configurations that influence the mechanisms through which performance feedback works and the resulting outcomes

We explored three contextual configurations:

- *Theory 7* Financial incentives and sanctions influence providers' responses to the feedback and public reporting of performance data.
- *Theory 8* The perceived credibility of performance data influences providers' responses to the feedback and public reporting of performance data.
- *Theory 9* The degree to which performance data are 'actionable' influences providers' responses to the feedback and public reporting of performance data.

Owing to the paucity of evidence evaluating the use of PROMs as a tool to improve patients care, much of the evidence to test theories 7 and 8 was drawn from the public reporting and feedback of other performance data. Providers perceive mandatory public reporting programmes initiated by regulators or national or state governments as being driven by political motives, while those initiated by employers or insurance companies were perceived as being driven by a desire to cut costs or increase a provider's market share. Under these circumstances, providers criticised these data as lacking credibility because they were based on data designed for a different purpose, had inadequate methods of case mix-adjustment and did not reflect what was clinically important. When these data were fed back privately, with no public reporting, providers either ignored these data or attempted to improve their data collection practices. When these data were publicly reported, they often led providers to focus on improving those areas of care subjected to measurement, at the expense of other areas of care: so-called 'tunnel vision' or 'effort substitution'. Under these circumstances, a provider's motivation for responding was to protect their professional reputation and to mitigate any damage that the reporting of their performance in the media had done. The addition of financial incentives, while also improving performance in measured areas in the short term, led to threshold effects and gaming, especially when the incentivised behaviours were not seen as supporting the care of patients.

Providers responded to these publicly reported data when their own internally collected data also suggested that there was a problem with the quality of care. Furthermore, externally mandated publicly reported data, particularly that focusing on outcomes, did not enable providers to identify the causes of any areas of poor care. Rather, additional investigations were needed, which required additional resources and the 'know-how' and capacity to carry them out.

Public reporting systems that are clinician-led were perceived as being driven by a desire to improve the quality of patient care, which secured greater clinical involvement. Clinician involvement in these programmes ensured that the indicators represented aspects of care that clinicians perceived as important, and clinical ownership of the sources of data, indicator specification and methods of case-mix adjustment meant it was difficult for them to dismiss or ignore these data. Under these circumstances, providers took steps to improve patient care in order to be as good as or better than their peers, and did so through sharing and learning from best practices. However, even under these conditions, action to improve patient care depended on the provider's experience of QI. Those providers with more experience of QI initiatives were more likely to make a greater number of sustainable improvements.

Providers valued data, such as patient experience data, that were timely, specific and provided a clear indication of which care processes needed to be improved. However, the feedback of these data did not always lead to providers taking steps to improve patient care. When change did occur, efforts were more likely to be directed at less complex, discrete organisational aspects of care. Changes that required clinicians to modify the interpersonal aspects of their care or that were more complex were perceived as more difficult to implement. Furthermore, changes to one aspect of patient care could have unintended effects on other aspects of care. Significant and sustained improvements in patient care in response to the feedback of performance data can be achieved only through system- and organisation-wide strategies.

Chapter 7: programme theories underlying the use of individual patient-reported outcome measures data to support the care of patients

Most PROMs have been psychometrically tested for use in research rather than routine clinical practice. Research has focused on the development, psychometric testing and application of these measures in research. There has been a proliferation in the number of PROMs, and a number of different types of PROMs exist, each with different strengths and weaknesses. However, the concepts measured by these PROMs, especially HRQoL, are contested.

The different functions of PROMs in the care of individual patients are:

- screening for patients' problems
- clinically monitoring patients' problems
- supporting patient-centred care.

Several models exist outlining the processes through which the feedback of PROMs data may improve the processes and outcomes of care. We identified a number of counter-theories challenging the idea that PROMs feedback may improve patient care, including:

- Collection of PROMs data is not practical or feasible.
- The psychometric properties or structure of PROMs do not support their use in clinical practice.
- PROMs may constrain rather than support the clinician–patient relationship.
- The challenges of using PROMs for multiple purposes.

We developed a logic model for the feedback of PROMs data in the care of individual patients, which formed the structure of our evidence synthesis.

Chapter 8: patient-reported outcome measures as a tool to support patients in raising or sharing concerns with clinicians

We tested and refined one programme theory and one counter-theory about the ways in which PROMs may support patients to raise or share their concerns with clinicians:

- *Theory 10* PROMs completion enables patients to share or raise issues with clinicians by acting as a reminder, facilitating self-reflection or giving patients 'permission' to raise issues.
- *Theory 11* PROMs completion and review may constrain the development of the patient–clinician relationship or hinder the flow of the consultation.

We hypothesised that a number of contextual configurations may influence these mechanisms and the resulting outcomes:

- *Theory 12* The structure and format of the PROM.
- *Theory 13* The existing nature of the patient–clinician relationship.
- *Theory 14* Whether PROMs completion is subjected to monetary rewards and sanctions, and whether PROMs data are also intended for use as an indicator of service quality.

Whether PROMs support or constrain patients in sharing or raising issues with clinicians depends on the structure of the PROM. Standardised PROMs were useful for those patients who preferred not to talk about personal or sensitive issues, as these helped them to share information. However, clinicians in primary and secondary mental health settings and palliative care perceived that standardised PROMs constrained the patient–clinician relationship because they trivialised patient emotions or did not capture the complex and dynamic nature of patients’ problems. They were also difficult to incorporate into the flow of consultations because the requirements to uphold the quantitative psychometric properties of the instrument restricted opportunities for sense-making or relationship building.

When there were no incentives attached to their use, clinicians avoided using standardised PROMs because they did not support their care of patients. In some situations, they also adapted or changed the PROMs to make them more useable, which may have compromised their validity as an instrument to support the care of individual patients. When clinicians were financially incentivised to use standardised PROMs, they used a wide range of tactics to both adapt the PROMs to fit with their interactions with patients but also to avoid being penalised for avoiding PROMs use. These strategies compromised the validity of the PROM both as a tool to support patient care and as an indicator of care quality.

Clinicians and patients perceived that individualised PROMs supported relationship-building when used in first assessments, because they enabled the patient to ‘tell their story’. However, individualised PROMs were less useful as an outcome measure to judge change over time, as the requirement to narrow the patients’ narrative into five cues could potentially distort patients’ experiences. Differences in the ways in which these cues were defined between patients, between patients and interviewers, and over time, challenged their use as a valid measure of change.

Chapter 9: patient-reported outcome measures as a tool for raising clinicians’ awareness of patients’ concerns

We began by testing three theories about the ways in which PROMs feedback may prompt discussion of patients’ concerns during the consultation and either directly or indirectly, through changes to patient management, improve patient well-being:

- *Theory 15* PROMs feedback increases the discussion of HRQoL issues during the consultation and this directly leads to improvements in patient well-being.
- *Theory 16* PROMs feedback increases the discussion of HRQoL issues during the consultation, which leads to changes to patient management that improve patient well-being.
- *Theory 17* PROMs feedback does not change clinicians’ communication practices during the consultation.

Patient-reported outcome measures feedback can increase discussion of symptoms in the consultation, and, in one study,³²⁷ this led directly to improvements in patient well-being. The mechanism underlying this process is that patients feel more comfortable raising both physical and psychosocial or non-medical issues with their doctors and are more likely to initiate discussion about these issues in the consultation. It is necessary that the clinician also receives the PROMs data to enable patients to feel comfortable raising issues, as this sends a further message to the patient that the doctor is interested in their concerns. PROMs feedback does not necessarily improve patient well-being through increased discussion leading to management changes. Clinicians rarely made changes to patient management as a result of PROMs feedback; changes to chemotherapy regimens were largely dependent on biomedical issues. Supportive treatment changes were made in response to issues that could directly be attributed to the patient’s cancer or its treatment.

Patient-reported outcome measures do not substantially change doctors’ communication practices during the consultation. Consultations still tended to focus on symptoms rather than psychosocial issues, and clinicians were not more likely to initiate discussions on these latter issues. One study indicated that common side effects of systemic treatment for cancer were discussed often, irrespective of whether

PROMs indicated they were a problem to the patient, while another found that the discussion of symptoms was dependent on the severity of the problem. However, the discussion of functional issues was not related to the severity of the problem. We then tested a further two theories to explain these findings:

- *Theory 18* Clinicians or patients do not consider it within their remit or judge that it is not appropriate to discuss psychosocial or functional issues during the consultation.
- *Theory 19* Clinicians avoid discussion of issues that they feel they cannot or do not know how to treat.

Doctors see their remit as dealing with issues specifically related to the patient's cancer and its treatment, and consider that it is the nurses' role to address wider psychosocial issues, a perception shared by nurses themselves. We found that doctors used a style of questioning in the consultation that focused on determining whether both specific and non-specific symptoms are related to cancer and its treatment. Doctors used a similar style of questioning when PROMs data were available, and closed down discussion of those symptoms that were either not a problem for the patient or unrelated to their cancer or its treatment. PROMs do not distinguish between 'problems related to cancer and its treatment' and 'other problems'; consequently, doctors did not take PROMs scores at face value but engaged in a process to understand the cause of the problem. This explains why PROMs do not shift the focus of the consultation to psychosocial issues; to do so would require a shift in doctors' perceptions of their remit. Professional groups' perceptions of their remit are socialised through many years of training and are mutually reinforced through the division of labour in the everyday practices of different clinical groups. PROMs feedback alone is not sufficient to change these practices, which are shaped by organisation-wide structures that both produce and reinforce these professional boundaries.

Clinicians were concerned that PROMs feedback may raise issues they are unable to address. In oncology settings, doctors closed down discussion of issues that they felt were not amenable to treatment. In palliative care settings, although nurses recognised that PROMs may raise issues that fall outside their remit and that they could not address, they recognised that 'just discussing' these issues could have therapeutic value for patients. These differences can be explained by a number of contextual factors, including the setting and professional group in question. In oncology, the focus is on improving patients' length of life at the same time as maintaining their quality of life; however, sometimes the impact of treatment on length of life may take precedence. Furthermore, doctors see their remit as dealing with the biomedical or physical aspects of cancer and that they are trained to 'fix things'. In palliative care, the focus of care is to maximise a patient's quality of life, and, although nurses are also trained to 'fix things', in this setting they also see it as their role to consider the wider psychosocial aspects of a patient's life.

Strengths and weaknesses of our review

We recognise that our review was necessarily focused on certain programme theories and that other important areas, such as the use of PROMs to promote patient self-management,²⁹¹ the use of electronic patient-reported outcome measures,²⁹³ the use of item banks³⁶¹ and the ways in which patients interpret and make use of quality information,²⁰⁰ were not covered in our report. Realist review is a flexible and iterative review methodology; this is both a strength and a weakness. We acknowledge that if another team conducted a realist review in this area, they might have selected different programme theories and thus tested different ideas and assumptions about how and in which circumstances PROMs feedback works. However, a comprehensive list of programme theories was identified and a range of stakeholders were involved in identifying these theories; therefore, we feel confident that they have face validity. We recognise that stakeholders were not as actively involved in prioritising and selecting the theories to review owing to the open nature of the discussion at our stakeholder workshop and patient group meetings. We would recommend that future reviews consider the use of more structured approaches, such as the nominal group technique,²³⁹ to involve stakeholders in theory selection.

We also acknowledge that we cannot rule out the possibility that an important study was missed from our review. However, we do not feel that these possibilities undermine the validity and value of our review. The aim of this review was not to conduct an exhaustive search for all studies, but rather to comprehensively sample these studies. A further strength of realist review is that the unit of analysis is not the intervention, but the programme theories that underpin it. For many of our theories relating to the feedback of aggregated PROMs data, there was a paucity of evidence specifically evaluating the use of PROMs data by providers. Instead of being unable to conduct a review, we were able to draw on studies that had evaluated different interventions that shared the same programme theory as PROMs feedback in order to test these theories. However, we recognise that it is important to draw attention to the different evidence bases used in the review; the review of aggregate PROMs data drew on a much wider range of interventions than the review of the individual PROMs data, which largely focused on studies that had specifically evaluated the use of PROMs in clinical practice.

We conducted our review following the RAMESES quality standards for realist reviews³⁶² and have reported our methodology and findings following the RAMESES realist review reporting standards.⁸³ We have provided a detailed account of our search methodologies and the process of theory elicitation and selection. Furthermore, providing a study-by-study account of each study has enabled the reader to see both how and why we came to our conclusions. We feel that these strategies have enhanced the transparency and thus improved the validity of our review. We also feel that the review has a number of valuable lessons for practice and research, which we discuss below.

Discussion and implications for practice

In this section we discuss the implications of our findings for practice and future research. We consider the findings of each review in turn, starting with our review of the feedback of aggregate level PROMs and performance data. We recognise that, for many of the theories tested in our review of aggregate PROMs, there was a paucity of literature specifically related to PROMs data. In realist synthesis, the unit of analysis is not the interventions but the programme theories that underpin them, and so we drew on evaluations of a wide range of performance data to test our theories. Thus, a focus on programme theories means that the findings from realist synthesis can offer a generalisable understanding of how contextual factors shape the mechanisms through which interventions sharing the same programme theories work. Here, we apply this learning specifically to the use of aggregate PROMs data.

Feedback of aggregate-level patient-reported outcome measures data

Our review found that patients rarely used performance data to inform their choice of hospital, and we have no reason to believe that PROMs data are an exception to this finding. Our review suggests that this is because patients prefer to rely on their own experience and that of friends and family when choosing providers. We acknowledge that there are also other reasons, not covered in our review, that patients may not use performance data; for example, they may have difficulty accessing and interpreting these data.²⁰⁰ It is possible that the new website 'My NHS'¹⁰⁶ will improve patients' access to and interpretation of these data. Our review identified that, in some circumstances, providers used performance data to improve the quality of patient care. We showed that a number of different contextual configurations influenced the degree to which they achieved their intended outcomes. These included the perceived credibility of data, which depended on perceived motivations of the body responsible for initiating the schemes, and on the source of data, indicator specification and case-mix adjustment; whether or not any financial incentives were attached; and the 'actionability' of the data, which depended on timeliness, specificity/level of aggregation and the extent to which the data clearly indicated the source of the problem.

Here, we identify how these configurations relate to three current public reporting of performance initiatives in the UK: the national PROMs programme, the national patient experience surveys conducted in hospital trusts and GP surgeries, and the FFT conducted in hospitals. *Table 6* summarises these characteristics.

TABLE 6 A comparison of the characteristics of different public reporting of performance initiatives in the English NHS

Contextual factor	English PROMs programme	Patient experience data	FFT
Credibility			
Who initiates	NHS England	GP survey – NHS England Other hospital inpatient and community surveys – CQC	NHS England
Source of data	Patient questionnaires linked to Hospital Episode Statistics	GP survey: random sample of patients registered with their GP for at least 6 months; hospital/ community surveys: patients discharged from the trust/service within the sampling month (and previous months if required to obtain necessary sample size)	Convenience sample of patients within 48 hours of discharge
Case-mix adjustment	Yes, developed by HSCIC; process available online	No	No
Financial incentives	Yes for participation (> 50%) and performance (health gain not significantly below national average at 99.8% significance level)	Not currently but GP patient survey items previously included in QOF Length of consultation Access	National CQUIN target 2014–15
Actionability			
Timeliness	Published annually, for previous year's data	Published annually for current year's data	Published monthly
Level of aggregation	Provider, but hospitals can request surgeon-level data	Provider and ward level	Provider and ward/service level
Type of data	Outcome/patient outcome	Process/patient experience	Process/patient experience

These programmes are examples of government-initiated performance feedback schemes that are publicly reported. However, *Table 6* shows that these programmes vary in terms of the contextual configurations we identified as important. Our review is focused on PROMs, and we now focus our discussion on the implications of our review for the use of PROMs data to improve patient care.

The use of financial incentives

We found that the impact of financial incentives was complex and depended on (1) whether the incentivised behaviour was seen by clinicians as indicative of what they would consider good patient care, and (2) the levels of care that clinicians were expected to demonstrate to receive the financial rewards. If performance indicators were not credible to clinicians but clinicians were incentivised to collect them, this could lead to the gaming or manipulation of performance data. Since 2014–15, providers have been incentivised for both their participation in and their performance on the indicators in the national PROMs programme. On the surface, it is difficult to see how providers could 'game' the collection of PROMs data in the same way that Mannion *et al.*'s¹⁸ respondents gamed the data collected under the performance assessment framework, as PROMs data rely on patients' responses. Current data on the HSCIC website indicate considerable variation in participation rates between providers, suggesting wide differences in whether and how providers administer questionnaires and manage the process of completion. Providers may selectively include patients who are likely to have better outcomes; however, this is hard to predict in advance, and the evidence suggests that people who currently respond to PROMs questionnaires are likely to have better baseline functioning in the first instance.^{67,68} Providers may refuse to administer PROMs questionnaires to patients altogether; however, evidence suggests that this may not work in their favour, as providers with lower participation rates are more likely to be erroneously classified as outliers.¹⁵³

Credibility

The Royal College of Surgeons was involved in the process of selecting the PROMs for the national PROMs programme.¹⁴ Both in this development phase and subsequently, considerable work was undertaken to explore the response rates and validity of these data for this purpose.^{67,68,107} However, as Devlin *et al.*¹²³ highlighted, the PROMs selected were originally developed for research purposes, and the intended functions of the PROMs programme are 'far removed from the initial intentions' of instrument developers. In our review, we found that clinicians remained sceptical about the validity of PROMs data for a number of reasons, including their reliance on patients' subjective reports and the process of case-mix adjustment.^{97,194} GPs were also not convinced that patient experience data reflected what they considered to be 'good care.'²²³ Indeed, recent evidence indicates that following good practice in terms of antibiotics prescribing may reduce patient satisfaction.³⁶³ These findings suggest that clinicians need further evidence of the validity of patients' subjective reports and their relevance to improving the quality of patient care. For example, the HSCIC has recently published a report outlining how different NHS trusts have used PROMs data to improve patient care.³⁶⁴ The HSCIC publish the methodology through which case-mix adjustment is calculated; however, it is unclear to what extent clinicians are involved in this process or whether they understand the basis on which case-mix adjustment is operationalised. Our findings would suggest that actively involving professional organisations in this case-mix adjustment process and explaining how this process is achieved in a language that clinicians can understand could improve the credibility of the adjustment process and PROMs data in the eyes of clinicians. Furthermore, our findings suggest that there is little variation between providers in their performance¹⁷³ within the national PROMs programme, and there may be other services and settings in which there is greater variation in practice that would benefit from the routine collection of PROMs data.

We also found that the perceived motivation behind the public reporting initiative was an important determinant of how providers responded to these data. In particular, public reporting programmes that were initiated by the government were often perceived as being driven by political motives rather than by a desire to improve patient care, and could risk leading to 'tunnel vision' in organisations with less experience of QI initiatives. We found some evidence that clinicians were concerned that the publication of PROMs data may be used for political ends.¹⁹⁴ Providers need more evidence that PROMs data can inform improvements in patient care. However, one of the biggest challenges faced by clinicians using PROMs data to improve patient care is identifying the source of the problem and implementing a solution, an issue we turn to next.

Actionability

We found that providers valued timely data that enabled them to clearly identify the problem. Our review identified two key challenges for the use of PROMs data for improving patient care. Currently, the time taken to collect PROMs data, link to Hospital Episode Statistics data, adjust for case mix and produce the indicators means that PROMs data reflect care provided 1 year ago and, as such, lack timeliness. Under the current methodology, it is difficult to see how the timeliness of PROMs data can be improved. One option is to collect PROMs data electronically, but this brings its own challenges in terms of response rates and engagement from key stakeholders.³⁶⁵

However, perhaps the most significant challenge to the use of PROMs data to improve patient care is the issue of attributability. Our review suggested that performance data, particularly those focused on outcomes, act as 'tin openers', not 'dials'. PROMs data can classify providers as 'statistical' outliers but this does not necessarily mean that those providers are providing poor care, nor does it inform providers of the possible causes of any poor care. Providers are expected to conduct further investigations to identify alternative explanations for their outlier status and, if none are found, explore the possible causes of their outlier status. Our review identified that providers could do this only if they routinely collected their own internal quality data and had the necessary infrastructure in place to embark on further investigations. Providers who were more experienced in QI activities were more likely to have these structures in place, although the direction of causation is not clear here.^{98,230} Therefore, providers need more support and guidance on how to collect their own internal data, how to rule out alternative explanations for their outlier status and how to explore the possible causes of their outlier status. However, we also found that even

when providers were given support to carry out this additional investigation, implementing changes remained a challenge.²³⁸ Bringing about significant, sustainable QI required a system-wide approach to change. The exploration of key factors involved in this process was beyond the scope of this review, but it has been explored in other reviews.^{89,366,367}

Furthermore, PROMs data are only one of many pieces of performance information that trust boards receive and have to address. When providers are not classed as outliers, the PROMs data run the risk of being ignored, as providers pay more attention to the measured areas of poor performance (so-called 'tunnel vision'). It could also be argued that greater improvements in patient care could be realised if all providers, not just those classed as negative outliers, used PROMs data to improve patient care. To support this and to avoid tunnel vision, further support and guidance are needed to enable providers to integrate and interpret PROMs data in the context of other pieces of performance information so that they can access a bigger picture of their performance. This will require effective co-ordination across different hospital departments, but also efficient lines of communication between the hospital board and staff working on the ground. In many of the studies included in our review, we found that performance data were poorly disseminated within the organisation. However, we would argue that dissemination should be a process with both a top-down and a bottom-up approach, so that staff on the ground are aware of and know how to interpret performance data but also so that staff can bring their insights about the possible causes of problems.

The feedback of patient-reported outcome measures data in the care of individual patients

Our review highlighted the importance of considering how PROMs feedback in the care of individual patients affects not only the information exchange and decision-making functions of the consultation but also the relationship-building function.³⁰² Other studies have also highlighted the tensions between implementing standardised approaches to patient assessment and care and maintaining and building a relationship with patients.^{241,284,285,368,369} Our review suggests that PROMs function more as a tool to support patients in raising issues with clinicians than they do in substantially changing clinicians' communication practices with patients. Patients valued both standardised and individualised PROMs as tools to do this, although our review also suggests that thought is required about which patients may benefit and which may not. Patients who find it difficult to raise sensitive issues verbally find standardised PROMs useful to share these issues with clinicians. However, for some patients, particularly those with terminal illnesses, PROMs completion can be an emotional experience and they may not wish to complete the measure. If they do complete it, they require support at the time of completion to manage this process.

Our review suggests that in settings such as palliative care and psychotherapy, clinicians viewed individualised PROMs as useful for building rapport and supporting the therapeutic process. However, individualised PROMs are more time-consuming to complete and it is not clear whether such measures would be feasible to collect in more time-pressured settings such as primary care or outpatient appointments for patients with long-term conditions. Our review suggests that, in these settings, nurses may be the most appropriate people to collect these data with patients, an issue we return to in the next sections. Furthermore, clinicians felt that individualised measures were not useful to measure change over time, which raises important questions about the role of individualised measures in patient care. Our review would suggest that they function as a 'conversation opener' rather than as an 'outcome measure'.

We found that an important reason that PROMs feedback did not substantially change clinicians' communication practices or focus discussion on psychosocial issues was that this would require a shift in clinicians' perceptions of their remit. Professional boundaries and remit are developed through the socialisation that different professional groups experience during training, and they are mutually reinforced by the division of labour among professional groups and the organisational structures that exist to support these ways of working. PROMs feedback alone is not going to change this. Patients with long-term conditions are cared for by many different clinicians, and these clinicians are often organised in multidisciplinary teams. Therefore, our review would suggest that the use of PROMs in the care of individual patients should focus on ensuring that the multiple different clinicians gain access to this information so that issues can be addressed by the clinicians

with the appropriate remit. This can best be achieved by integrating PROMs collection into the patients' electronic record.^{291,294} However, in order to work, this approach would require a clear division of labour and line of responsibility for which professional group should respond to and address different PROM items and subscale, as has been found in other studies.³¹¹

Collection of patient-reported outcome measures for multiple purposes?

Finally, our review highlights the difficulties of collecting PROMs data both to support individual patient care and as a measures of service performance or quality.¹⁶ Van der Wees²⁶⁹ identified three models of PROMs data collection internationally and we have provided examples of each model below:

1. the collection of PROMs data by national, regional or state organisations for the purposes of measuring service quality (e.g. the England national PROMs programme¹⁷³)
2. the collection of PROMs data by individual providers for the care of individual patients (e.g. the collection of PROMs in oncology^{326,327})
3. data collected by individual providers for use in the care of individual patients that is then aggregated for use as a measure of service quality (e.g. the use of standardised depression questionnaires in the QOF⁸¹ and the collection of PROMs in the IAPT programme^{339,340,370}).

In the first model, PROMs are used purely as a measure of service quality and our review of PROMs feedback at the aggregate level examined the circumstances in which and processes through which this may improve patient care. However, in this model, opportunities for the use of PROMs to inform the care of individual patients are lost, as data on individual patients are not fed back to individual clinicians. In the second model, the focus is on the care of individual patients, and one could argue that opportunities are lost for data collected to inform service-level improvements. In the third model, PROMs data are collected for multiple functions. However, the multiple stakeholders involved in this model have different and sometimes competing drivers for collecting these data. Our review of the aggregate feedback of PROMs data found that the perceived underlying driver of the reporting system played a significant role in providers' engagement with and response to the programme. Providers were much more engaged if the programme focused on improving the quality of patient care and measured things that mattered to clinicians. Our review of individual PROMs data also found that unless the data collected supports clinicians in their care of individual patients, clinicians will either avoid using the PROMs or adapt the use of PROMs in a way that may undermine their value for one or both purposes. This raises two important challenges for this integrated model. The first is whether or not it is possible to achieve the necessary trust between different stakeholders such that the collection of PROMs feedback is perceived as being focused on improving the quality of patient care.²⁶⁹ The second is whether or not the same PROM can be used for multiple purposes, as PROMs that clinicians find useful in assessment of patients, such as individualised measures, are not useful as indicators of service quality.¹⁶ We would argue that these two questions remain a challenge and should be the focus of future research.

Recommendations for future research

In many ways, the research agenda for understanding and improving the impact of PROMs is part of a wider agenda of the impact of all rigorous quantitative performance data (such as infection rates, patient incident rates, morbidity or disability outcome measures). There may be some additional issues with PROMs, given the reluctance of some clinicians to give as much credence and importance to patients' views as to their own objective biomedical measures.

The rather modest impact of PROMs on clinical behaviour and practice is partly a consequence of clinicians and managers of providers not knowing how to interpret the data they receive and how to use it to explore possible reasons for any suboptimal performance. Research is needed to identify the needs of clinicians and managers in assisting them to make use of PROMs output.

There are areas of health care in which our understanding of how best to incorporate PROMs is unclear. One area is emergency admissions, in which an unexpected event occurs in an otherwise healthy person (even if they have a long-term condition that puts them at risk of such an event). A pre-event PROM is clearly not feasible, so other approaches to determining their outcome is needed. Research is needed to determine if and how PROMs can be usefully employed in such situations. Another area is when a patient's cognitive function is deteriorating over time, most obviously relevant in patients with dementia. In addition, the impact of such a condition on the HRQoL of a patient's relatives and friends may be as great or even greater. Research is needed as to how PROMs can best be used both to assist individual clinical and social management and to compare the performance of providers.

Our review also suggested that clinicians value qualitative narrative feedback (e.g. comment cards and complaints that provide insight into patient experiences). However, this information is difficult to summarise and feed back to clinicians in way that can be read and understood quickly. Further research is needed to explore ways of feeding these data back to clinicians and to combine and incorporate them with quantitative data in a busy clinical setting.

Our review identified that there are many more studies that have evaluated and explored PROMs feedback in the care of individual patients than studies that have examined their use at an aggregate level to improve patient care. There are few published studies that have explored in detail how differently performing providers (e.g. positive and negative and at average levels) have used PROMs to identify the areas of care that could be improved and how they have used PROMs data to guide improvements. There is also a need to understand how information from aggregated PROMs data is integrated with other performance data and used to inform QI activities. We would argue that such studies should take a mixed-methods approach and combine a qualitative analysis of the process of QI with a quantitative analysis of impacts. There are also few published studies that have explored how providers have collected PROMs data to support both the care of individual patients and service QI. It would be useful to learn from organisations about the process through which this was achieved and, again, mixed-methods studies would be valuable.

As the health and social care systems are brought together and, under the Five Year Forward View, integrated models of care focused on local health economies emerge, the role and use of PROMs could prove helpful at both an individual and an aggregate level. Research is needed to explore how PROMs could best contribute to the changing care landscape. Specifically, there is a need for the development of IT systems, software and integration of PROMs in electronic health records and evaluation of the practicalities and feasibility of implementation.

Research is also needed to explore the potential benefit of PROMs data to support patients' choice of health care, specifically at the publicly available aggregate level. Furthermore, further exploration is required into the usefulness of PROMs at an individual patient level as a mechanism for self-management. Offering patients the opportunity to complete PROMs online or incorporating PROMs completion into interventions to support self-management is a growing area of research. Advocates argue that patient use of PROMs to self-monitor their condition can improve patient activation and empowerment and, in turn, lead to improve self-management.^{287,371} However, others³⁰⁹ have cautioned that digital self-monitoring may require patients to engage in such monitoring at particular times of the day; thus, 'empowerment' becomes a 'set of obligations'. Our review has also shown that not all patients welcome or benefit from PROMs completion, so careful thought is required to decide who this intervention can be offered to. Evaluative research in this area has found a mixed picture in terms of the impact of patient monitoring on self-management and patient outcomes.^{313,318,372-375} Mixed-methods research is needed to explore the contexts in which and mechanisms through which these interventions achieve their intended outcomes.

Finally, our review suggests the idea that PROMs feedback can provide the impetus for different groups of health professionals to discuss and then address aspects of patients' health (e.g. psychosocial issues) may be thwarted because different professional groups vary in the extent to which they see this as part of their clinical remit. Future research examining PROMs feedback in the care of individual patients should explicitly

explore how PROMs feedback challenges or may be constrained by interprofessional boundaries and perceptions of clinical remit. Research is also needed to explore what it is about the culture of clinical regimes and the content of different medical curricula that operates as a formative influence in the partial response to PROMs. *Table 7* provides a summary of these recommendations in priority order, along with an indication of possible methodologies and methods for addressing them.

TABLE 7 Future research priorities

Priority	Research recommendation	Possible methodologies/methods
1	How do providers with different levels of performance respond to PROMs feedback and how are these data integrated with other forms of performance data?	Realist evaluation comparative case studies Ethnographic observations and interviews Controlled interrupted time series
2	How can PROMs successfully be used to support individual patient care and monitor service performance?	Realist evaluation comparative case studies Ethnographic observations and interviews Controlled interrupted time series
3	What are the needs of clinicians and managers to assist them in making use of PROMs feedback?	Qualitative interviews Surveys Controlled user experiments
4	The use of PROMs data to support patient self-management and the integration of PROMs data with patient information systems	Realist evaluation Comparative case studies Ethnographic observations and interviews
5	How PROMs data and other performance data are used to support patient choice	Qualitative interviews Controlled user experiments
6	How PROMs feedback in the care of individual patients is supported or constrained by professional boundaries and the role of professional training in use of PROMs data	Ethnographic observations and interviews
7	Exploration of new areas for collection and use of aggregated PROMs data, such as emergency care and dementia care	Theory-driven literature review to scope possible methods Feasibility testing of different methods

Acknowledgements

Contributions of authors

All authors read and provided feedback on drafts of the report. All authors attended project team meetings and contributed to the intellectual content of the report.

Joanne Greenhalgh (Associate Professor, Realist Synthesis and PROMs Research) was principal investigator, designed the study, wrote the protocol and takes overall responsibility for the quality of the study. She wrote the first draft and then finalised all chapters in the final report, with contribution to designing the searches, study screening and selection, data extraction and synthesis.

Sonia Dalkin (Research Fellow, Psychology, Realist Methods) contributed to study screening and selection, data extraction and synthesis, and the writing of *Chapters 3, 4, 5 and 7*.

Kate Gooding (Research Fellow, Realist Methods, International Development) contributed to study screening and selection, data extraction and synthesis, the writing of *Chapters 8, 9 and 10*, and proofreading the final report.

Elizabeth Gibbons (Senior Research Scientist, PROMs, Patient Experience, Nursing) contributed to study screening and selection, data extraction and synthesis, and the writing of *Chapters 7 and 10*.

Judy Wright (Senior Information Specialist, Literature Searching and Management) contributed to literature searching and the writing of *Chapters 2 and 6* on methodology.

David Meads (Lecturer, Health Economics) contributed to the writing of *Chapter 4*.

Nick Black (Professor of Health Services Research, PROMs and the National PROMs Programme, Quality of Health Care, Assessment of Healthcare Providers) contributed to the writing of *Chapter 3*.

Jose Maria Valderas (Professor of Health Services Research and Policy, PROMs, Management of Long-Term Conditions, Primary Care) contributed to the writing of *Chapter 7*.

Ray Pawson (Emeritus Professor of Social Research Methodology, Realist Synthesis) contributed to methodological guidance throughout the synthesis and to the writing of *Chapters 4, 5, 8 and 9*.

Acknowledged contributors

Project team members

Liz Lingard provided advice and feedback on the aggregate level synthesis.

Jane Jackson provided advice and feedback on the aggregate level synthesis.

Laurence Wood provided PPI perspective on the synthesis and chaired the PPI group.

Patient group members

Eileen Exeter.

Rosie Hassaman.

Gill Riley.

We would also like to thank the editor and anonymous reviewers for their useful and constructive comments on the first draft of the report.

Publication

Greenhalgh J, Pawson R, Wright J, Black N, Valderas JM, Meads D, *et al.* Functionality and feedback: a protocol for a realist synthesis of the collation, interpretation and utilisation of PROMs data to improve patient care. *BMJ Open* 2014;**4**.

Data sharing statement

All documents and papers on which this review is based are available in the public domain. Summaries of all studies included in this report are provided within the report.

References

1. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010;**340**:c186. <http://dx.doi.org/10.1136/bmj.c186>
2. Guyatt GH, Veldhuyzen Van Zanten SJ, Feeny DH, Patrick DL. Measuring quality of life in clinical trials: a taxonomy and review. *CMAJ* 1989;**140**:1441–8.
3. Black N, Browne J, van der Meulen J, Jamieson L, Copley L, Lewsey J. Is there overutilisation of cataract surgery in England? *Br J Ophthalmol* 2009;**93**:13–17. <http://dx.doi.org/10.1136/bjo.2007.136150>
4. Neuburger J, Hutchings A, Allwood D, Black N, van der Meulen JH. Sociodemographic differences in the severity and duration of disease amongst patients undergoing hip or knee replacement surgery. *J Public Health* 2012;**34**:421–9. <http://dx.doi.org/10.1093/pubmed/fdr119>
5. Department of Health. *High Quality Care for All: NHS Next Stage Review Final Report*. London: Department of Health; 2008.
6. Department of Health. *Equity and Excellence: Liberating the NHS*. London: Department of Health; 2010.
7. Francis R. *Report of the Mid Staffordshire NHS Foundation Trust Public Enquiry*. London: The Stationery Office; 2013.
8. Nuffield Trust. *Rating Providers for Quality: A Policy Worth Pursuing?* London: Nuffield Trust; 2013.
9. Haywood KL, Garratt AM, Carrivick S, Mangnall J, Skevington SM. Continence specialists use of quality of life information in routine practice: a national survey of practitioners. *Qual Life Res* 2009;**18**:423–33. <http://dx.doi.org/10.1007/s11136-009-9459-1>
10. Bausewein C, Simon ST, Benalia H, Downing J, Mwangi-Powell FN, Daveson BA, et al. Implementing patient reported outcome measures (PROMs) in palliative care – users’ cry for help. *Health Qual Life Outcomes* 2011;**9**:27. <http://dx.doi.org/10.1186/1477-7525-9-27>
11. Snyder CF, Aaronson NK, Choucrair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res* 2012;**21**:1305–14. <http://dx.doi.org/10.1007/s11136-011-0054-x>
12. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res* 2009;**18**:115–23. <http://dx.doi.org/10.1007/s11136-008-9430-6>
13. Higginson IJ, Carr AJ. Measuring quality of life: using quality of life measures in the clinical setting. *BMJ* 2001;**322**:1297–300. <http://dx.doi.org/10.1136/bmj.322.7297.1297>
14. Browne J, Jamieson E, Lewsey J, Van der Meulen J, Black N, Cairns J, et al. *Patient Reported Outcome Measures (PROMs) in Elective Surgery: A Report to the Department of Health*. London: London School of Hygiene and Tropical Medicine; 2007.
15. Lewis JM. The politics and consequences of performance measurement. *Policy Soc* 2015;**34**:1–12. <http://dx.doi.org/10.1016/j.polsoc.2015.03.001>
16. Wolpert M. Uses and abuses of patient reported outcome measures (PROMs): potential iatrogenic impact of PROMs implementation and how it can be mitigated. *Adm Policy Ment Health* 2014;**41**:141–5. <http://dx.doi.org/10.1007/s10488-013-0509-1>

17. Paterson C, Thomas K, Manasse A, Cooke H, Peace G. Measure Yourself Concerns and Wellbeing (MYCaW): an individualised questionnaire for evaluating outcome in cancer support care that includes complementary therapies. *Complement Ther Med* 2007;**15**:38–45. <http://dx.doi.org/10.1016/j.ctim.2006.03.006>
18. Mannion R, Davies H, Marshall M. Impact of star performance ratings in English acute hospital trusts. *J Health Serv Res Policy* 2005;**10**:18–24. <http://dx.doi.org/10.1258/1355819052801877>
19. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med* 2008;**148**:111–23. <http://dx.doi.org/10.7326/0003-4819-148-2-200801150-00006>
20. Berwick DM. The science of improvement. *JAMA* 2008;**299**:1182–4. <http://dx.doi.org/10.1001/jama.299.10.1182>
21. Pawson R. Evidence and policy and naming and shaming. *Policy Stud* 2002;**23**:211–30. <http://dx.doi.org/10.1080/0144287022000045993>
22. Harrison S. New labour, modernisation and the medical labour process. *J Soc Policy* 2002;**31**:465–85. <http://dx.doi.org/10.1017/S0047279402006694>
23. Haywood K, Marshall S, Fitzpatrick R. Patient participation in the consultation process: a structured review of intervention strategies. *Patient Educ Couns* 2006;**63**:12–23. <http://dx.doi.org/10.1016/j.pec.2005.10.005>
24. Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Qual Life Res* 2013;**22**:2265–78. <http://dx.doi.org/10.1007/s11136-013-0390-0>
25. Weingarten SR, Kim CS, Stone EG, Kristopaitis RJ, Pelter M, Sandhu M. Can peer-comparison feedback improve patient functional status? *Am J Manag Care* 2000;**6**:35–9.
26. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. *JAMA* 2000;**283**:1866–74. <http://dx.doi.org/10.1001/jama.283.14.1866>
27. Totten AM, Wagner J, Tiwari A, O'Haire C, Griffin J, Walker M. Closing the quality gap: revisiting the state of the science (vol. 5: public reporting as a quality improvement strategy). *Evid Rep Technol Assess* 2012;**208.5**:1–645.
28. Ketelaar NA, Faber MJ, Flottorp S, Rygh LH, Deane KH, Eccles MP. Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations. *Cochrane Database Syst Rev* 2011;**11**:CD004538. <http://dx.doi.org/10.1002/14651858.cd004538.pub2>
29. Hannan EL, Kumar D, Racz M, Siu AL, Chassin MR. New York State's cardiac surgery reporting system: four years later. *Ann Thorac Surg* 1994;**58**:1852–7. [http://dx.doi.org/10.1016/0003-4975\(94\)91726-4](http://dx.doi.org/10.1016/0003-4975(94)91726-4)
30. Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA* 1994;**271**:761–6. <http://dx.doi.org/10.1001/jama.1994.03510340051033>
31. Cutler DM, Huckman RS, Labndrum MB. The role of information in medical markets: an analysis of publicly reported outcomes in cardiac surgery. *AER* 2004;**94**:342–6. <http://dx.doi.org/10.1257/0002828041301993>
32. Jha AK, Epstein AM. The predictive accuracy of the New York State coronary artery bypass surgery report-card system. *Health Aff* 2006;**25**:844–55. <http://dx.doi.org/10.1377/hlthaff.25.3.844>

33. Peterson ED, DeLong ER, Jollis JG, Muhlbaier LH, Mark DB. The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly. *J Am Coll Cardiol* 1998;**32**:993–9. [http://dx.doi.org/10.1016/S0735-1097\(98\)00332-5](http://dx.doi.org/10.1016/S0735-1097(98)00332-5)
34. Carey JS, Danielsen B, Junod FL, Rossiter SJ, Stabile BE. The California Cardiac Surgery and Intervention Project: evolution of a public reporting program. *Am Surg* 2006;**72**:978–83.
35. Bridgewater B, Grayson AD, Brooks N, Grotte G, Fabri BM, Au J, et al. Has the publication of cardiac surgery outcome data been associated with changes in practice in northwest England: an analysis of 25,730 patients undergoing CABG surgery under 30 surgeons over eight years. *Heart* 2007;**93**:744–8. <http://dx.doi.org/10.1136/hrt.2006.106393>
36. Hannan EL, Sarrazin MS, Doran DR, Rosenthal GE. Provider profiling and quality improvement efforts in coronary artery bypass graft surgery: the effect on short-term mortality among Medicare beneficiaries. *Med Care* 2003;**41**:1164–72. <http://dx.doi.org/10.1097/01.MLR.0000088452.82637.40>
37. Moscucci M, Eagle KA, Share D, Smith D, De Franco AC, O'Donnell M, et al. Public reporting and case selection for percutaneous coronary interventions: an analysis from two large multicenter percutaneous coronary intervention databases. *J Am Coll Cardiol* 2005;**45**:1759–65. <http://dx.doi.org/10.1016/j.jacc.2005.01.055>
38. Romano PS, Marcin JP, Dai JJ, Yang XD, Kravitz RL, Rocke DM, et al. Impact of public reporting of coronary artery bypass graft surgery performance data on market share, mortality, and patient selection. *Med Care* 2011;**49**:1118–25. <http://dx.doi.org/10.1097/MLR.0b013e3182358c78>
39. Guru V, Fremes SE, Naylor CD, Austin PC, Shrive FM, Ghali WA, Tu JV, Cardiac Care Network of Ontario. Public versus private institutional performance reporting: what is mandatory for quality improvement? *Am Heart J* 2006;**152**:573–8. <http://dx.doi.org/10.1016/j.ahj.2005.10.026>
40. Ghali WA, Ash AS, Hall RE, Moskowitz MA. Statewide quality improvement initiatives and mortality after cardiac surgery. *JAMA* 1997;**277**:379–82. <http://dx.doi.org/10.1001/jama.1997.03540290031028>
41. Krägeloh CU, Czuba KJ, Billington DR, Kersten P, Siegert RJ. Using feedback from patient-reported outcome measures in mental health services: a scoping study and typology. *Psychiatr Serv* 2015;**66**:224–41. <http://dx.doi.org/10.1176/appi.ps.201400141>
42. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, MacGillivray S. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *J Clin Oncol* 2014;**32**:1480–501. <http://dx.doi.org/10.1200/JCO.2013.53.5948>
43. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res* 2013;**13**:211. <http://dx.doi.org/10.1186/1472-6963-13-211>
44. Carlier IV, Meuldijk D, van Vliet IM, van Fenema EM, van der Wee NJ, Zitman FG. [Empirical evidence for the effectiveness of routine outcome monitoring. A study of the literature.] *Tijdschr Psychiatr* 2012;**54**:121–8.
45. Carlier IV, Meuldijk D, Van Vliet IM, Van Fenema E, Van der Wee NJ, Zitman FG. Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract* 2012;**18**:104–10. <http://dx.doi.org/10.1111/j.1365-2753.2010.01543.x>
46. Luckett T, Butow PN, King MT. Improving patient outcomes through the routine use of patient-reported data in cancer clinics: future directions. *Psychooncology* 2009;**18**:1129–38. <http://dx.doi.org/10.1002/pon.1545>

47. Knaup C, Koesters M, Schoefer D, Becker T, Puschner B. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *Br J Psychiatry* 2009;**195**:15–22. <http://dx.doi.org/10.1192/bjp.bp.108.053967>
48. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, *et al*. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res* 2008;**17**:179–93. <http://dx.doi.org/10.1007/s11136-007-9295-0>
49. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract* 2006;**12**:559–68. <http://dx.doi.org/10.1111/j.1365-2753.2006.00650.x>
50. Lambert MJ, Whipple JL, Hawkins EJ, Vermeersch DA, Neilson SL, Smart DW. Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clin Psychol* 2003;**10**:288–301. <http://dx.doi.org/10.1093/clipsy.bpg025>
51. Gilbody SM, House AO, Sheldon TA. Routinely administered questionnaires for depression and anxiety: systematic review. *BMJ* 2001;**322**:406–9. <http://dx.doi.org/10.1136/bmj.322.7283.406>
52. Gilbody SM, House AO, Sheldon T. Routine administration of Health Related Quality of Life (HRQoL) and needs assessment instruments to improve psychological outcome – a systematic review. *Psychol Med* 2002;**32**:1345–56. <http://dx.doi.org/10.1017/S0033291702006001>
53. Espallargues M, Valderas JM, Alonso J. Provision of feedback on perceived health status to health care professionals: a systematic review of its impact. *Med Care* 2000;**38**:175–86. <http://dx.doi.org/10.1097/00005650-200002000-00007>
54. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *J Eval Clin Pract* 1999;**5**:401–16. <http://dx.doi.org/10.1046/j.1365-2753.1999.00209.x>
55. Alsaleh K. Routine administration of standardized questionnaires that assess aspects of patients' quality of life in medical oncology clinics: a systematic review. *J Egypt Natl Canc Inst* 2013;**25**:63–70. <http://dx.doi.org/10.1016/j.jnci.2013.03.001>
56. Mitchell AJ. Screening for cancer-related distress: when is implementation successful and when is it unsuccessful? *Acta Oncol* 2013;**52**:216–24. <http://dx.doi.org/10.3109/0284186X.2012.745949>
57. Gonçalves Bradley DC, Gibbons C, Ricci-Cabello I, Bobrovitz NJH, Gibbons EJ, Kotzeva A, *et al*. Routine provision of information on patient-reported outcome measures to healthcare providers and patients in clinical practice. *Cochrane Database Syst Rev* 2015;**4**:CD011589. <http://dx.doi.org/10.1002/14651858.cd011589>
58. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf* 2014;**23**:508–18. <http://dx.doi.org/10.1136/bmjqs-2013-002524>
59. Antunes B, Harding R, Higginson IJ, EUROIMPACT. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. *Palliat Med* 2014;**28**:158–75. <http://dx.doi.org/10.1177/0269216313491619>
60. Etkind SN, Daveson BA, Kwok W, Witt J, Bausewein C, Higginson IJ, Murtagh FE. Capture, transfer, and feedback of patient-centered outcomes data in palliative care populations: does it make a difference? A systematic review. *J Pain Symptom Manage* 2015;**49**:611–24. <http://dx.doi.org/10.1016/j.jpainsymman.2014.07.010>
61. Duncan EA, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. *BMC Health Serv Res* 2012;**12**:96. <http://dx.doi.org/10.1186/1472-6963-12-96>

62. Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? *Soc Sci Med* 2005;**60**:833–43. <http://dx.doi.org/10.1016/j.socscimed.2004.06.022>
63. Takeuchi EE, Keding A, Awad N, Hofmann U, Campbell LJ, Selby PJ, *et al.* Impact of patient-reported outcomes in oncology: a longitudinal analysis of patient-physician communication. *J Clin Oncol* 2011;**29**:2910–17. <http://dx.doi.org/10.1200/JCO.2010.32.2453>
64. Detmar SB, Muller MJ, Wever LD, Schornagel JH, Aaronson NK. The patient-physician relationship. Patient-physician communication during outpatient palliative treatment visits: an observational study. *JAMA* 2001;**285**:1351–7. <http://dx.doi.org/10.1001/jama.285.10.1351>
65. Greenhalgh J, Abhyankar P, McCluskey S, Takeuchi E, Velikova G. How do doctors refer to patient-reported outcome measures (PROMS) in oncology consultations? *Qual Life Res* 2013;**22**:939–50. <http://dx.doi.org/10.1007/s11136-012-0218-3>
66. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;**346**:f167. <http://dx.doi.org/10.1136/bmj.f167>
67. Hutchings A, Neuburger J, Grosse Frie K, Black N, van der Meulen J. Factors associated with non-response in routine use of patient reported outcome measures after elective surgery in England. *Health Qual Life Outcomes* 2012;**10**:34. <http://dx.doi.org/10.1186/1477-7525-10-34>
68. Hutchings A, Grosse Frie K, Neuburger J, van der Meulen J, Black N. Late response to patient-reported outcome questionnaires after surgery was associated with worse outcome. *J Clin Epidemiol* 2013;**66**:218–25. <http://dx.doi.org/10.1016/j.jclinepi.2012.09.001>
69. Davies HT, Crombie IK. Interpreting health outcomes. *J Eval Clin Pract* 1997;**3**:187–99. <http://dx.doi.org/10.1046/j.1365-2753.1997.00003.x>
70. Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007;**335**:648–50. <http://dx.doi.org/10.1136/bmj.39317.641296.AD>
71. Hildon Z, Neuburger J, Allwood D, van der Meulen J, Black N. Clinicians' and patients' views of metrics of change derived from patient reported outcome measures (PROMs) for comparing providers' performance of surgery. *BMC Health Serv Res* 2012;**12**:171. <http://dx.doi.org/10.1186/1472-6963-12-171>
72. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of the Oxford hip and knee scores. *J Bone Joint Surg Br* 2007;**89**:1010–14. <http://dx.doi.org/10.1302/0301-620X.89B8.19424>
73. Valderas JM, Fitzpatrick R, Roland M. Using health status to measure NHS performance: another step into the dark for the health reform in England. *BMJ Qual Saf* 2012;**21**:352–3. <http://dx.doi.org/10.1136/bmjqs-2011-000184>
74. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA* 2005;**293**:1239–44. <http://dx.doi.org/10.1001/jama.293.10.1239>
75. Hafner JM, Williams SC, Koss RG, Tschurtz BA, Schmaltz SP, Loeb JM. The perceived impact of public reporting hospital performance data: interviews with hospital staff. *Int J Qual Health Care* 2011;**23**:697–704. <http://dx.doi.org/10.1093/intqhc/mzr056>
76. West D. Unneeded surgery may be costing the NHS millions. *Health Service Journal*, 14 May 2009.
77. Marshall M, McLoughlin V. How do patients use information on health providers? *BMJ* 2010;**341**:c5272. <http://dx.doi.org/10.1136/bmj.c5272>
78. Magee H, Davis LJ, Coulter A. Public views on healthcare performance indicators and patient choice. *J R Soc Med* 2003;**96**:338–42. <http://dx.doi.org/10.1258/jrsm.96.7.338>

79. Dixon A, Robertson R, Appleby J, Burge P, Devlin NJ, Magee H. *Patient Choice: How Patients Choose and How Providers Respond*. London: The King's Fund; 2010.
80. Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract* 2011;**61**:e279–89. <http://dx.doi.org/10.3399/bjgp11X572472>
81. Dowrick C, Leydon GM, McBride A, Howe A, Burgess H, Clarke P, *et al*. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ* 2009;**338**:b663. <http://dx.doi.org/10.1136/bmj.b663>
82. Pawson R. *Evidence Based Policy: A Realist Perspective*. London: Sage; 2006. <http://dx.doi.org/10.4135/9781849209120>
83. Wong CK, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist synthesis. *BMC Med* 2013;**11**:21. <http://dx.doi.org/10.1186/1741-7015-11-21>
84. Killoran A, Kelly M. *Evidence-based Public Health: Effectiveness and Efficiency*. Oxford: Oxford University Press; 2009. <http://dx.doi.org/10.1093/acprof:oso/9780199563623.001.0001>
85. Pawson R, Tilley N. *Realistic Evaluation*. London: Sage; 1997.
86. Merton R. *On Theoretical Sociology: Five Essays, Old and New*. New York, NY: Free Press; 1967.
87. Pawson R. Digging for nuggets: how bad research can yield good evidence. *Int J Soc Res Methodol* 2006;**9**:127–42. <http://dx.doi.org/10.1080/13645570600595314>
88. Greenhalgh J, Pawson R, Wright J, Black N, Valderas JM, Meads D, *et al*. Functionality and feedback: a protocol for a realist synthesis of the collation, interpretation and utilisation of PROMs data to improve patient care. *BMJ Open* 2014;**4**:e005601. <http://dx.doi.org/10.1136/bmjopen-2014-005601>
89. Pawson R, Greenhalgh J, Brennan C, Glidewell E. Do reviews of healthcare interventions teach us how to improve healthcare systems? *Soc Sci Med* 2014;**114**:129–37. <http://dx.doi.org/10.1016/j.socscimed.2014.05.032>
90. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006;**2**:CD000259. <http://dx.doi.org/10.1002/14651858.cd000259.pub2>
91. Wolfram Cox JR, Mann L, Samson D. Benchmarking as a mixed metaphor: disentangling assumptions of competition and collaboration. *J Manage Stud* 1997;**34**:285–314. <http://dx.doi.org/10.1111/1467-6486.00052>
92. Northcott D, Lewellyn S. Benchmarking in UK Health: a gap between policy and practice? *Benchmarking* 2005;**12**:419–35. <http://dx.doi.org/10.1108/14635770510619357>
93. van Helden GV, Tillema S. In search of a benchmarking theory for the public sector. *Financ Accountability Manage* 2005;**21**:337–61. <http://dx.doi.org/10.1111/j.0267-4424.2005.00224.x>
94. Chang L. Managerial responses to externally imposed performance measurement in the NHS: a institutional theory perspective. *Financ Accountability Manage* 2006;**22**:63–85. <http://dx.doi.org/10.1111/j.0267-4424.2006.00393.x>
95. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;**6**:CD000259. <http://dx.doi.org/10.1002/14651858.cd000259.pub3>
96. Flottorp SA, Jamtvedt G, Gibis B, McKee M. *Using Audit and Feedback to Health Professionals to Improve the Quality and Safety of Health Care*. Copenhagen: World Health Organization; 2010.

97. Boyce MB, Browne JP, Greenhalgh J. Surgeon's experiences of receiving peer benchmarked feedback using patient-reported outcome measures: a qualitative study. *Implement Sci* 2014;**9**:84. <http://dx.doi.org/10.1186/1748-5908-9-84>
98. Davies HT. Public release of performance data and quality improvement: internal responses to external data by US health care providers. *Qual Health Care* 2001;**10**:104–10. <http://dx.doi.org/10.1136/qhc.10.2.104>
99. Exworthy M, Wilkinson EK, McColl A, Moore M, Roderick P, Smith H, Gabbay J. The role of performance indicators in changing the autonomy of the general practice profession in the UK. *Soc Sci Med* 2003;**56**:1493–504. [http://dx.doi.org/10.1016/S0277-9536\(02\)00151-X](http://dx.doi.org/10.1016/S0277-9536(02)00151-X)
100. Smith PC. Performance measurement in health care: history, challenges and prospects. *Public Money Management* 2005;**25**:213–20. <http://dx.doi.org/10.1080/09540962.2005.10600096>
101. Department of Health. *The New NHS: Modern, Dependable*. London: Department of Health; 1997.
102. Department of Health. *The NHS Plan*. London: Department of Health; 2000.
103. Great Britain. *Health and Social Care Act*. London: The Stationery Office; 2008.
104. Department of Health. *The NHS Outcomes Framework 2011/12*. London: Department of Health; 2010.
105. Kennedy I. *The Report of the Public Inquiry into Children's Heart Surgery at the Bristol Royal Infirmary 1984–95: Learning from Bristol*. London: The Stationery Office; 2001.
106. NHS England. *My NHS*. 2015. URL: www.nhs.uk/service-search/performance/search (accessed 10 December 2015).
107. Smith S, Cano S, Lamping DL, Staniszewska S, Browne J, Lewsey J, et al. *Patient-Reported Outcome Measures (PROMs) for Routine Use in Treatment Centres: Recommendations Based on a Review of the Scientific Evidence*. London: London School of Hygiene and Tropical Medicine/Royal College of Nursing; 2005.
108. Department of Health. *Patient Reported Outcome Measures in England. The Case-Mix Adjustment Methodology*. London: Department of Health; 2012.
109. Department of Health. *Patient Reported Outcome Measures in England: A Methodology for Identifying Potential Outliers*. London: Department of Health; 2012.
110. Bridgewater B, Keogh B. Surgical 'league tables': ischaemic heart disease. *Heart* 2008;**94**:936–42. <http://dx.doi.org/10.1136/hrt.2008.143602>
111. Bridgewater B, Hickey GL, Cooper G, Deanfield J, Roxburgh J, Society for Cardiothoracic Surgery in Great Britain and Ireland. Publishing cardiac surgery mortality rates: lessons for other specialties. *BMJ* 2013;**346**:f1139. <http://dx.doi.org/10.1136/bmj.f1139>
112. Department of Health. *Mental Health Information Strategy*. London: Department of Health; 2001.
113. National IAPT Programme Team. *The IAPT Data Handbook: Guidance on Recording and Monitoring Outcomes to Support Local Evidence-Based Practice Version*. London: National IAPT Programme Team; 2011.
114. Anekwe L. QOF depression indicators face axe. *Pulse*, 9 June 2011.
115. Stokes T. *QOF Changes are Ground-Breaking and Will Need Monitoring*. Chesham: Guidelines in Practice; 2014. URL: www.guidelinesinpractice.co.uk/apr_14_stokes_general_apr14#.VbjTGPIVikp (accessed 17 November 2016).
116. McDonald R, Zaidi S, Todd S, Konteh F, Hussain K, Roe J, et al. *A Qualitative and Quantitative Evaluation of the Introduction of Best Practice Tariffs*. London: Department of Health; 2012.

117. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. *Med Care* 2003;**41**(Suppl. 1):I30–8. <http://dx.doi.org/10.1097/00005650-200301001-00004>
118. Hibbard JH, Stockard J, Tusler M. Hospital performance reports: impact on quality, market share, and reputation. *Health Aff* 2005;**24**:1150–60. <http://dx.doi.org/10.1377/hlthaff.24.4.1150>
119. Department of Health. *Guidance on the Routine Collection of Patient Reported Outcome Measures (PROMs)*. London: Department of Health; 2008.
120. Health and Social Care Information Centre. *Monthly Patient Reported Outcome Measures (PROMs) in England*. London: HSCIC; 2015.
121. Marshall MN, Shekelle PG, Leatherman S, Brook RH. Public disclosure of performance data: learning from the US experience. *Qual Health Care* 2000;**9**:53–7. <http://dx.doi.org/10.1136/qhc.9.1.53>
122. Pinto DS, Pride YB. Paved with good intentions and marred by half-truths. *J Am Coll Cardiol* 2013;**62**:416–17. <http://dx.doi.org/10.1016/j.jacc.2013.04.026>
123. Devlin NJ, Appleby J, Buxton M, Vallance-Owen A. *Getting the Most Out of PROMs: Putting Health Outcomes at the Heart of Decision Making*. London: The King's Fund; 2010.
124. Vallance-Owen A, Cubbin S, Warren V, Matthews B. Outcome monitoring to facilitate clinical governance; experience from a national programme in the independent sector. *J Public Health* 2004;**26**:187–92. <http://dx.doi.org/10.1093/pubmed/fdh127>
125. Timmins N. NHS goes to the PROMS. *BMJ* 2008;**336**:1464–5. <http://dx.doi.org/10.1136/bmj.39618.627951.80>
126. Rosen HM, Green BA. The HCFA excess mortality lists: a methodological critique. *Hosp Health Serv Adm* 1987;**32**:119–27.
127. Bridgewater B. Why doctors' outcomes should be published in the press. *BMJ* 2005;**331**:1210. <http://dx.doi.org/10.1136/bmj.331.7526.1210>
128. Sofaer S, Hibbard JH. *Best Practices in Public Report No 3: How to Maximise Public Awareness and Use of Comparative Quality Reports through Effective Promotion and Dissemination Strategies*. Rockville, MD: Agency for Healthcare Research and Quality; 2010.
129. Mebius C. Public disclosure of health care performance reports: comments and reflections from Sweden. *Int J Qual Health Care* 1999;**11**:102–3. <http://dx.doi.org/10.1093/intqhc/11.2.102>
130. Devlin NJ, Parkin D, Browne J. *Using the EQ-5D as a Performance Measurement Tool in the NHS*. Report No. 09/03. London: Department of Economics, City University London; 2009.
131. Burge P, Devlin NJ, Appleby J, Rohr C, Grant J. Do patients always prefer quicker treatment? A discrete choice analysis of patients' stated preferences in the London Patient Choice Project. *Appl Health Econ Health Policy* 2004;**3**:183–94. <http://dx.doi.org/10.2165/00148365-200403040-00002>
132. Kahneman D, Knetsch JL, Thaler RH. Anomalies: the endowment effect, loss aversion, and status quo bias. *J Econ Perspect* 1991;**5**:193–206. <http://dx.doi.org/10.1257/jep.5.1.193>
133. Gabe J, Exworthy M, Jones IR, Smith G. Towards a sociology of disclosure: the case of surgical performance. *Sociology Compass* 2012;**6**:908–22. <http://dx.doi.org/10.1111/j.1751-9020.2012.00490.x>
134. Kullgren JT, Werner RM. Counterpoint: will public reporting of health-care quality measures inform and educate patients? No. *Chest* 2011;**140**:1117–20. <http://dx.doi.org/10.1378/chest.11-2094>

135. Fotaki M. Patient choice and equity in the British National Health Service: towards developing an alternative framework. *Sociol Health Illn* 2010;**32**:898–913. <http://dx.doi.org/10.1111/j.1467-9566.2010.01254.x>
136. West D. Exclusive: patient choice is not key to improving performance, says Hunt. *Health Service Journal*, 26 November 2014.
137. Shaw J, Taylor R, Dix K. *Uses and Abuses for Performance Data in Healthcare*. London: Dr Foster; 2015.
138. Smith P. On the unintended consequences of publishing performance data in the public sector. *Int J Public Admin* 1995;**18**:277–310. <http://dx.doi.org/10.1080/01900699508525011>
139. Braithwaite J, Mannion R. Government plans for public reporting of performance data in health care: the case against. *Med J Aust* 2011;**195**:41.
140. Kelman S, Friedman JN. Performance improvement and performance dysfunction: an empirical examination of distortionary impacts on the emergency room wait-time target in the English National Health Service. *J Public Admin Res Theory* 2009;**19**:917–46. <http://dx.doi.org/10.1093/jopart/mun028>
141. Holmstrom B, Milgrom P. Multitask principal-agent analysis: incentive contracts, asset ownership and job design. *J Law Econ Organ* 1991;**7**:24–52. http://dx.doi.org/10.1093/jleo/7.special_issue.24
142. Bevan G, Hood C. What's measured is what matters: targets and gaming in the English public health care system. *Public Admin* 2006;**84**:517–38. <http://dx.doi.org/10.1111/j.1467-9299.2006.00600.x>
143. Bloor M. A minor office: the variable and socially constructed character of death certification in a Scottish city. *J Health Soc Behav* 1991;**32**:273–87. <http://dx.doi.org/10.2307/2136808>
144. Dixon-Woods M, Leslie M, Bion J, Tarrant C. What counts? An ethnographic study of infection data reported to a patient safety program. *Milbank Q* 2012;**90**:548–91. <http://dx.doi.org/10.1111/j.1468-0009.2012.00674.x>
145. British Medical Association (BMA). *BMA Warns Against Simplistic Surgeon Tables*. BMA; 2012. URL: <http://bma.org.uk/news-views-analysis/news/2012/december/bma-warns-against-simplistic-surgeon-tables> (accessed 6 July 2015).
146. Naylor CD. Public profiling of clinical performance. *JAMA* 2002;**287**:1323–5. <http://dx.doi.org/10.1001/jama.287.10.1323>
147. Marshall MN, Brook RH. Public reporting of comparative information about quality of healthcare. *Med J Aust* 2002;**176**:205–6.
148. Iezzoni LI, Ash AS, Schwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health* 1996;**86**:1379–87. <http://dx.doi.org/10.2105/AJPH.86.10.1379>
149. Grant SW, Grayson AD, Jackson M, Au J, Fabri BM, Grotte G, *et al*. Does the choice of risk-adjustment model influence the outcome of surgeon-specific mortality analysis? A retrospective analysis of 14,637 patients under 31 surgeons. *Heart* 2008;**94**:1044–9. <http://dx.doi.org/10.1136/hrt.2006.110478>
150. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;**363**:1147–54. [http://dx.doi.org/10.1016/S0140-6736\(04\)15901-1](http://dx.doi.org/10.1016/S0140-6736(04)15901-1)

151. Browne JP, Bastaki H, Dawson J. What is the optimal time point to assess patient-reported recovery after hip and knee replacement? A systematic review and analysis of routinely reported outcome data from the English patient-reported outcome measures programme. *Health Qual Life Outcomes* 2013;**11**:128. <http://dx.doi.org/10.1186/1477-7525-11-128>
152. Maynard A, Bloor K. Patient reported outcome measurement: learning to walk before we run. *J R Soc Med* 2010;**103**:129–32. <http://dx.doi.org/10.1258/jrsm.2010.090477>
153. Hutchings A, Neuburger J, van der Meulen J, Black N. Estimating recruitment rates for routine use of patient reported outcome measures and the impact on provider comparisons. *BMC Health Serv Res* 2014;**14**:66. <http://dx.doi.org/10.1186/1472-6963-14-66>
154. WHO Regional Office for Europe's Health Evidence Network. *How can Hospital Performance be Measured and Monitored?* Copenhagen: World Health Organization; 2003.
155. Guerriere M. Determining the utility of public reporting – too early to judge. *Healthc Pap* 2005;**6**:62–7. <http://dx.doi.org/10.12927/hcpap..17764>
156. NHS England. *The Friends and Family Test*. London: NHS England; 2014. URL: www.nhs.uk/NHSEngland/AboutNHSservices/Documents/FFTGuide_Final_1807_FINAL.pdf
157. Black N, Jenkinson C. Measuring patients' experiences and outcomes. *BMJ* 2009;**339**:b2495. <http://dx.doi.org/10.1136/bmj.b2495>
158. Hawkes N. Patients' rating of treatment tells you more about patients than hospitals, research concludes. *BMJ* 2013;**347**:f6916. <http://dx.doi.org/10.1136/bmj.f6916>
159. Carter N, Day P, Klein R. *How Organisations Measure Success: The Use of Performance Indicators in Government*. London: Routledge; 1995.
160. Carlisle D. NHS PROMs have proved effective in assessing patient care. *The Guardian*, 17 November 2010.
161. Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care* 2001;**13**:475–80. <http://dx.doi.org/10.1093/intqhc/13.6.475>
162. Long AF, Fairfield G. Confusion of levels in monitoring outcomes and/or process. *Lancet* 1996;**347**:1572. [http://dx.doi.org/10.1016/S0140-6736\(96\)91072-7](http://dx.doi.org/10.1016/S0140-6736(96)91072-7)
163. Powell AE, Davies HT, Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Qual Saf Health Care* 2003;**12**:122–8. <http://dx.doi.org/10.1136/qhc.12.2.122>
164. Bird SM, Cox D, Farewell VT, Goldstein H, Holt T, Smith PC. Performance indicators: good, bad, ugly. *J R Stat Soc A Sta* 2005;**168**:1–27. <http://dx.doi.org/10.1111/j.1467-985X.2004.00333.x>
165. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;**10**:407–15. [http://dx.doi.org/10.1016/0197-2456\(89\)90005-6](http://dx.doi.org/10.1016/0197-2456(89)90005-6)
166. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. *Psychol Bull* 1982;**92**:111–35. <http://dx.doi.org/10.1037/0033-2909.92.1.111>
167. Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. *Soc Sci Med* 2010;**70**:1618–25. <http://dx.doi.org/10.1016/j.socscimed.2010.01.039>
168. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 1996;**119**:254–84. <http://dx.doi.org/10.1037/0033-2909.119.2.254>

169. Meade PH. *A Guide to Benchmarking*. Perth, WA: Curtin University of Technology; 2007.
170. Department of Health. *Essence of Care*. London: Department of Health; 2010.
171. Oliver C. Strategic responses to institutional processes. *Acad Manag Rev* 1991;**16**:145–79. <http://dx.doi.org/10.5465/AMR.1991.4279002>
172. Exworthy M, Smith G, Gabe J, Jones IR. Disclosing clinical performance: the case of cardiac surgery. *J Health Organ Manag* 2010;**24**:571–83. <http://dx.doi.org/10.1108/14777261011088665>
173. Varagunam M, Hutchings A, Neuburger J, Black N. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. *J Health Serv Res Policy* 2014;**19**:77–84. <http://dx.doi.org/10.1177/1355819613506187>
174. Boyce MB, Browne JP. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient reported outcome measures – result from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study. *BMJ Open* 2015;**5**:e008325. <http://dx.doi.org/10.1136/bmjopen-2015-008325>
175. Bevan G, Skellern M. Does competition between hospitals improve clinical quality? A review of evidence from two eras of competition in the English NHS. *BMJ* 2011;**343**:d6470. <http://dx.doi.org/10.1136/bmj.d6470>
176. van der Veer SN, de Keizer NF, Ravelli ACJ, Tenkink S, Jager KJ. Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers. *Int J Med Inform* 2010;**79**:305–23. <http://dx.doi.org/10.1016/j.ijmedinf.2010.01.011>
177. Taylor A, Neuburger J, Walker K, Cromwell D, Groene O. How is feedback from national clinical audits used? Views from English National Health Service trust audit leads. *J Health Serv Res Policy* 2016;**21**:91–100. <http://dx.doi.org/10.1177/1355819615612826>
178. Royal College of Physicians. *Improving Lung Cancer Outcomes Project – ILCOP*. 2012. URL: www.rcplondon.ac.uk/projects/improving-lung-cancer-outcomes-project-ilcop (accessed 5 April 2016).
179. Hibbard JH, Stockard J, Tusler M. Does publicizing hospital performance stimulate quality improvement efforts? *Health Aff* 2003;**22**:84–94. <http://dx.doi.org/10.1377/hlthaff.22.2.84>
180. Foreman S, Kenkel D, Shea D. Cost and quality information and health care market reform. *Adv Health Econ Health Serv Res* 1995;**15**:137–53.
181. Wang J, Hockenberry J, Chou SY, Yang M. Do bad report cards have consequences? Impacts of publicly reported provider quality information on the CABG market in Pennsylvania. *J Health Econ* 2011;**30**:392–407. <http://dx.doi.org/10.1016/j.jhealeco.2010.11.006>
182. Mukamel DB, Mushlin AI. Quality of care information makes a difference: an analysis of market share and price changes after publication of the New York State Cardiac Surgery Mortality Reports. *Med Care* 1998;**36**:945–54. <http://dx.doi.org/10.1097/00005650-199807000-00002>
183. Romano PS, Zhou H. Do well-publicized risk-adjusted outcomes reports affect hospital volume? *Med Care* 2004;**42**:367–77. <http://dx.doi.org/10.1097/01.mlr.0000118872.33251.11>
184. Dranove D, Sfekas A. Start spreading the news: a structural estimate of the effects of New York hospital report cards. *J Health Econ* 2008;**27**:1201–7. <http://dx.doi.org/10.1016/j.jhealeco.2008.03.001>
185. Menemeyer ST, Morrissey MA, Howard LZ. Death and reputation: how consumers acted upon HCFA mortality information. *Inquiry* 1997;**34**:117–28.

186. Vladeck BC, Goodwin EJ, Myers LP, Sinisi M. Consumers and hospital use: the HCFA 'death list'. *Health Aff* 1988;**7**:122–5. <http://dx.doi.org/10.1377/hlthaff.7.1.122>
187. Howard DH, Kaplan B. Do report cards influence hospital choice? The case of kidney transplantation. *Inquiry* 2006;**43**:150–9. http://dx.doi.org/10.5034/inquiryjrnl_43.2.150
188. Baker DW, Einstadter D, Thomas C, Husak S, Gordon NH, Cebul RD. The effect of publicly reporting hospital performance on market share and risk-adjusted mortality at high-mortality hospitals. *Med Care* 2003;**41**:729–40. <http://dx.doi.org/10.1097/01.MLR.0000064640.66138.9A>
189. Wubker A, Auerland D, Wubker A. *Does Better Information about Hospital Quality Affect Patients' Choice? Empirical Findings from Germany*. Social Science Research Network; 2008.
190. Pope DG. Reacting to rankings: evidence from 'America's Best Hospitals'. *J Health Econ* 2009;**28**:1154–65. <http://dx.doi.org/10.1016/j.jhealeco.2009.08.006>
191. Chassin MR. Achieving and sustaining improved quality: lessons from New York State and cardiac surgery. *Health Aff* 2002;**21**:40–51. <http://dx.doi.org/10.1377/hlthaff.21.4.40>
192. Tu JV, Cameron C. Impact of an acute myocardial infarction report card in Ontario, Canada. *Int J Qual Health Care* 2003;**15**:131–7. <http://dx.doi.org/10.1093/intqhc/mzg015>
193. Guru V, Naylor CD, Fremes SE, Teoh K, Tu JV. Publicly reported provider outcomes: the concerns of cardiac surgeons in a single-payer system. *Can J Cardiol* 2009;**25**:33–8. [http://dx.doi.org/10.1016/S0828-282X\(09\)70020-0](http://dx.doi.org/10.1016/S0828-282X(09)70020-0)
194. Hildon Z, Allwood D, Black N. Patients' and clinicians' views of comparing the performance of providers of surgery: a qualitative study. *Health Expect* 2015;**18**:366–78. <http://dx.doi.org/10.1111/hex.12037>
195. Mehrotra A, Bodenheimer T, Dudley RA. Employers' efforts to measure and improve hospital quality: determinants of success. *Health Aff* 2003;**22**:60–71. <http://dx.doi.org/10.1377/hlthaff.22.2.60>
196. Greener I, Mannion R. Patient choice in the NHS: what is the effect of patient choice policies on patients and relationships in health economies. *Public Money Management* 2009;**29**:95–100. <http://dx.doi.org/10.1080/09540960902767972>
197. Berwick DM, Wald DL. Hospital leaders' opinions of the HCFA mortality data. *JAMA* 1990;**263**:247–9. <http://dx.doi.org/10.1001/jama.1990.03440020081037>
198. Rudd J, Glanz K. A survey of newspaper coverage of HCFA hospital mortality data. *Public Health Rep* 1991;**106**:517–23.
199. Greer AL. *Embracing Accountability: Physician Leadership, Public Reporting, and Teamwork in the Wisconsin Collaborative for Healthcare Quality*. New York, NY: The Commonwealth Fund; 2008.
200. Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consumers use quality-of-care information? A systematic review. *Med Care* 2009;**47**:1–8. <http://dx.doi.org/10.1097/MLR.0b013e3181808bb5>
201. Harris KM. Can high quality overcome consumer resistance to restricted provider access? Evidence from a health plan choice experiment. *Health Serv Res* 2002;**37**:551–71. <http://dx.doi.org/10.1111/1475-6773.00037>
202. Uhrig JD, Short PF. Testing the effect of quality reports on the health plan choices of Medicare beneficiaries. *Inquiry* 2002–3;**39**:355–71. http://dx.doi.org/10.5034/inquiryjrnl_39.4.355
203. Schoenbaum M, Spranca M, Elliott M, Bhattacharya J, Short PF. Health plan choice and information about out-of-pocket costs: an experimental analysis. *Inquiry* 2001;**38**:35–48. http://dx.doi.org/10.5034/inquiryjrnl_38.1.35

204. Spranca M, Kanouse DE, Elliott M, Short PF, Farley DO, Hays RD. Do consumer reports of health plan quality affect health plan selection? *Health Serv Res* 2000;**35**:933–47.
205. Hibbard JH, Berkman N, McCormack LA, Jael E. The impact of a CAHPS report on employee knowledge, beliefs, and decisions. *Med Care Res Rev* 2002;**59**:104–16. <http://dx.doi.org/10.1177/1077558702059001006>
206. Farley DO, Elliott MN, Short PF, Damiano P, Kanouse DE, Hays RD. Effect of CAHPS performance information on health plan choices by Iowa Medicaid beneficiaries. *Med Care Res Rev* 2002;**59**:319–36. <http://dx.doi.org/10.1177/107755870205900305>
207. Farley DO, Short PF, Elliott MN, Kanouse DE, Brown JA, Hays RD. Effects of CAHPS health plan performance information on plan choices by New Jersey Medicaid beneficiaries. *Health Serv Res* 2002;**37**:985–1007. <http://dx.doi.org/10.1034/j.1600-0560.2002.62.x>
208. Knutson DJ, Kind EA, Fowles JB, Adlis S. Impact of report cards on employees: a natural experiment. *Health Care Financ Rev* 1998;**20**:5–27.
209. Dixon S. *Report on the National Patient Choice Survey: February 2010 England*. London: Department of Health; 2010.
210. Abraham J, Sick B, Anderson J, Berg A, Dehmer C, Tufano A. Selecting a provider: what factors influence patients' decision making? *J Healthc Manag* 2011;**56**:99–114.
211. Dijs-Elsinga J, Otten W, Versluijs MM, Smeets HJ, Kievit J, Vree R, et al. Choosing a hospital for surgery: the importance of information on quality of care. *Med Decis Making* 2010;**30**:544–55. <http://dx.doi.org/10.1177/0272989X09357474>
212. Marang-van de Mheen PJ, Dijs-Elsinga J, Otten W, Versluijs M, Smeets HJ, Vree R, et al. The relative importance of quality of care information when choosing a hospital for surgical treatment: a hospital choice experiment. *Med Decis Making* 2011;**31**:816–27. <http://dx.doi.org/10.1177/0272989X10386799>
213. Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E, Beyer FR. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane Database Syst Rev* 2011;**7**:CD009255. <http://dx.doi.org/10.1002/14651858.cd009255>
214. Markovitz AA, Ryan AM. Pay-for-performance: disappointing results or masked heterogeneity. *Med Care Res Rev* 2017;**74**:3–78. <http://dx.doi.org/10.1177/1077558715619282>
215. Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 2007;**356**:486–96. <http://dx.doi.org/10.1056/NEJMs064964>
216. Friedberg MW, SteelFisher GK, Karp M, Schneider EC. Physician groups' use of data from patient experience surveys. *J Gen Intern Med* 2011;**26**:498–504. <http://dx.doi.org/10.1007/s11606-010-1597-1>
217. Alexander JA, Maeng D, Casalino LP, Rittenhouse D. Use of care management practices in small- and medium-sized physician groups: do public reporting of physician quality and financial incentives matter? *Health Serv Res* 2013;**48**:376–97. <http://dx.doi.org/10.1111/j.1475-6773.2012.01454.x>
218. Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, Salisbury C, Reeves D. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ* 2011;**342**:d3590. <http://dx.doi.org/10.1136/bmj.d3590>

219. Wilkinson EK, McColl A, Exworthy M, Roderick P, Smith H, Moore M, Gabbay J. Reactions to the use of evidence-based performance indicators in primary care: a qualitative study. *Qual Health Care* 2000;**9**:166–74. <http://dx.doi.org/10.1136/qhc.9.3.166>
220. Mannion R, Goddard M. Impact of published clinical outcomes data: case study in NHS hospital trusts. *BMJ* 2001;**323**:260–3. <http://dx.doi.org/10.1136/bmj.323.7307.260>
221. Mannion R, Goddard M. Public disclosure of comparative clinical performance data: lessons from the Scottish experience. *J Eval Clin Pract* 2003;**9**:277–86. <http://dx.doi.org/10.1046/j.1365-2753.2003.00388.x>
222. Elliott MN, Lehrman WG, Goldstein EH, Giordano LA, Beckett MK, Cohea CW, Cleary PD. Hospital survey shows improvements in patient experience. *Health Aff* 2010;**29**:2061–7. <http://dx.doi.org/10.1377/hlthaff.2009.0876>
223. Asprey A, Campbell JL, Newbould J, Cohn S, Carter M, Davey A, Roland M. Challenges to the credibility of patient feedback in primary healthcare settings: a qualitative study. *Br J Gen Pract* 2013;**63**:e200–8. <http://dx.doi.org/10.3399/bjgp13X664252>
224. Bradley EH, Holmboe ES, Mattera JA, Roumanis SA, Radford MJ, Krumholz HM. Data feedback efforts in quality improvement: lessons learned from US hospitals. *Qual Saf Health Care* 2004;**13**:26–31. <http://dx.doi.org/10.1136/qhc.13.1.26>
225. Luce JM, Thiel GD, Holland MR, Swig L, Currin SA, Luft HS. Use of risk-adjusted outcome data for quality improvement by public hospitals. *West J Med* 1996;**164**:410–14.
226. Rainwater JA, Romano PS, Antonius DM. The California Hospital Outcomes Project: how useful is California's report card for quality improvement? *Jt Comm J Qual Improv* 1998;**24**:31–9.
227. Romano PS, Rainwater JA, Antonius D. Grading the graders: how hospitals in California and New York perceive and interpret their report cards. *Med Care* 1999;**37**:295–305. <http://dx.doi.org/10.1097/00005650-199903000-00009>
228. Pham HH, Coughlan J, O'Malley AS. The impact of quality-reporting programs on hospital operations. *Health Aff* 2006;**25**:1412–22. <http://dx.doi.org/10.1377/hlthaff.25.5.1412>
229. Lamb GC, Smith MA, Weeks WB, Queram C. Publicly reported quality-of-care measures influenced Wisconsin physician groups to improve performance. *Health Aff* 2013;**32**:536–43. <http://dx.doi.org/10.1377/hlthaff.2012.1275>
230. Smith MA, Wright A, Queram C, Lamb GC. Public reporting helped drive quality improvement in outpatient diabetes care among Wisconsin physician groups. *Health Aff* 2012;**31**:570–7. <http://dx.doi.org/10.1377/hlthaff.2011.0853>
231. Vingerhoets E, Wensing M, Grol R. Feedback of patients' evaluations of general practice care: a randomised trial. *Qual Health Care* 2001;**10**:224–8. <http://dx.doi.org/10.1136/qhc.0100224>
232. Barr JK, Giannotti TE, Sofaer S, Duquette CE, Waters WJ, Petrillo MK. Using public reports of patient satisfaction for hospital quality improvement. *Health Serv Res* 2006;**41**:663–82. <http://dx.doi.org/10.1111/j.1475-6773.2006.00508.x>
233. Boyer L, Francois P, Doutre E, Weil G, Labarere J. Perception and use of the results of patient satisfaction surveys by care providers in a French teaching hospital. *Int J Qual Health Care* 2006;**18**:359–64. <http://dx.doi.org/10.1093/intqhc/mzl029>
234. Geissler KH, Friedberg MW, SteelFisher GK, Schneider EC. Motivators and barriers to using patient experience reports for performance improvement. *Med Care Res Rev* 2013;**70**:621–35. <http://dx.doi.org/10.1177/1077558713496319>

235. Reeves R, Seccombe I. Do patient surveys work? The influence of a national survey programme on local quality-improvement initiatives. *Qual Saf Health Care* 2008;**17**:437–41. <http://dx.doi.org/10.1136/qshc.2007.022749>
236. Boiko O, Campbell JL, Elmore N, Davey AF, Roland M, Burt J. The role of patient experience surveys in quality assurance and improvement: a focus group study in English general practice. *Health Expect* 2015;**18**:1982–94. <http://dx.doi.org/10.1111/hex.12298>
237. Reeves R, West E, Barron D. Facilitated patient experience feedback can improve nursing care: a pilot study for a phase III cluster randomised controlled trial. *BMC Health Serv Res* 2013;**13**:259. <http://dx.doi.org/10.1186/1472-6963-13-259>
238. Davies E, Shaller D, Edgman-Levitan S, Safran DG, Oftedahl G, Sakowski J, Cleary PD. Evaluating the use of a modified CAHPS survey to support improvements in patient-centred care: lessons from a quality improvement collaborative. *Health Expect* 2008;**11**:160–76. <http://dx.doi.org/10.1111/j.1369-7625.2007.00483.x>
239. Delbecq AL, van de Ven AH, Gustafson DH. *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. Glenview, IL: Scott, Foresman; 1975.
240. Van De A, Delbecq AL. Nominal versus interacting group processes for committee decision-making effectiveness. *Academy of Management Journal* 1971;**14**:203–12. <http://dx.doi.org/10.2307/255307>
241. Cowley S, Mitcheson J, Houston AM. Structuring health needs assessments: the medicalisation of health visiting. *Sociol Health Illn* 2004;**26**:503–26. <http://dx.doi.org/10.1111/j.0141-9889.2004.00403.x>
242. Slater A, Freeman E. Is the Palliative Care Outcome Scale useful to staff in a day hospice unit? *Int J Palliat Nurs* 2005;**11**:346–54. <http://dx.doi.org/10.12968/ijpn.2005.11.7.18487>
243. Kettis-Lindblad A, Ring L, Widmark E, Bendtsen P, Glimelius B. Patients' and doctors' views of using the schedule for individual quality of life in clinical practice. *J Support Oncol* 2007;**5**:281–7.
244. Annells M, Koch T. 'The real stuff': implications for nursing of assessing and measuring a terminally ill person's quality of life. *J Clin Nurs* 2001;**10**:806–12. <http://dx.doi.org/10.1046/j.1365-2702.2001.00546.x>
245. Guyatt GH, Ferrans CE, Halyard MY, Revicki DA, Symonds TL, Varricchio CG, *et al*. Exploration of the value of health-related quality-of-life information from clinical research and into clinical practice. *Mayo Clin Proc* 2007;**82**:1229–39. <http://dx.doi.org/10.4065/82.10.1229>
246. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993;**329**:977–86. <http://dx.doi.org/10.1056/NEJM199309303291401>
247. Pocock SJ. A perspective on the role of quality-of-life assessment in clinical trials. *Control Clin Trials* 1991;**12**(Suppl. 4):257–65. [http://dx.doi.org/10.1016/S0197-2456\(05\)80029-7](http://dx.doi.org/10.1016/S0197-2456(05)80029-7)
248. Spilker B, Revicki D. Taxonomy of Quality of Life. In Spilker B, editor. *Quality of Life and Pharmacoeconomics in Clinical Trials*. Philadelphia, PA: Lippincott-Raven Publishers; 1996. pp. 25–31.
249. Hunt SM. The problem of quality of life. *Qual Life Res* 1997;**6**:205–12.
250. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;**273**:59–65. <http://dx.doi.org/10.1001/jama.1995.03520250075037>
251. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ* 2002;**324**:1417. <http://dx.doi.org/10.1136/bmj.324.7351.1417>

252. Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 Health Survey: Manual and Interpretation Guide*. Boston, MA: The Health Institute, New England Medical Center; 1993.
253. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;**33**:337–43. <http://dx.doi.org/10.3109/07853890109002087>
254. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br* 1998;**80**:63–9. <http://dx.doi.org/10.1302/0301-620X.80B1.7859>
255. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, *et al*. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;**85**:365–76. <http://dx.doi.org/10.1093/jnci/85.5.365>
256. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, *et al*. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;**11**:570–9.
257. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992;**45**:1341–5. [http://dx.doi.org/10.1016/0895-4356\(92\)90194-R](http://dx.doi.org/10.1016/0895-4356(92)90194-R)
258. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, Stein RE. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;**11**:193–205. <http://dx.doi.org/10.1023/A:1015291021312>
259. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;**19**:539–49. <http://dx.doi.org/10.1007/s11136-010-9606-8>
260. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, *et al*. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;**22**:1889–905. <http://dx.doi.org/10.1007/s11136-012-0344-y>
261. Greenhalgh J, Long AF, Brett AJ, Grant MJ. Reviewing and selecting outcome measures for use in routine practice. *J Eval Clin Pract* 1998;**4**:339–50. <http://dx.doi.org/10.1111/j.1365-2753.1998.tb00097.x>
262. Wright JG. Evaluating the outcome of treatment. Shouldn't we be asking patients if they are better? *J Clin Epidemiol* 2000;**53**:549–53. [http://dx.doi.org/10.1016/s0895-4356\(99\)00225-5](http://dx.doi.org/10.1016/s0895-4356(99)00225-5)
263. Staniszewska S, Haywood KL, Brett J, Tutton L. Patient and public involvement in patient-reported outcome measures: evolution not revolution. *Patient* 2012;**5**:79–87. <http://dx.doi.org/10.2165/11597150-000000000-00000>
264. O'Boyle C, McGee H, Hickey A, Joyce CRB, Browne J, O'Malley KJ, *et al*. *The Schedule for the Evaluation of Individual Quality of Life (SEIQoL): Administration Manual*. Dublin: Royal College of Surgeons of Ireland; 1993.
265. Paterson C. Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. *BMJ* 1996;**312**:1016–20. <http://dx.doi.org/10.1136/bmj.312.7037.1016>
266. Carr AJ. Margaret Holroyd Prize Essay. A patient-centred approach to evaluation and treatment in rheumatoid arthritis: the development of a clinical tool to measure patient-perceived handicap. *Br J Rheumatol* 1996;**35**:921–32. <http://dx.doi.org/10.1093/rheumatology/35.10.921>

267. Ruta DA, Garratt AM, Leng M, Russell IT, MacDonald LM. A new approach to the measurement of quality of life. The Patient-Generated Index. *Med Care* 1994;**32**:1109–26. <http://dx.doi.org/10.1097/00005650-199411000-00004>
268. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity – establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1 – eliciting concepts for a new PRO instrument. *Value Health* 2011;**14**:967–77. <http://dx.doi.org/10.1016/j.jval.2011.06.014>
269. Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *Milbank Q* 2014;**92**:754–75. <http://dx.doi.org/10.1111/1468-0009.12091>
270. Deyo RA, Carter WB. Strategies for improving and expanding the application of health status measures in clinical settings. A researcher-developer viewpoint. *Med Care* 1992;**30**(Suppl. 5):176–86. <http://dx.doi.org/10.1097/00005650-199205001-00015>
271. Wagner AK, Vickrey BG. The routine use of health-related quality of life measures in the care of patients with epilepsy: rationale and research agenda. *Qual Life Res* 1995;**4**:169–77. <http://dx.doi.org/10.1007/BF01833610>
272. Frost MH, Bonomi AE, Cappelleri JC, Schünemann HJ, Moynihan TJ, Aaronson NK, Clinical Significance Consensus Meeting Group. Applying quality-of-life data formally and systematically into clinical practice. *Mayo Clin Proc* 2007;**82**:1214–28. <http://dx.doi.org/10.4065/82.10.1214>
273. Ebrahim S. Clinical and public health perspectives and applications of health-related quality of life measurement. *Soc Sci Med* 1995;**41**:1383–94. [http://dx.doi.org/10.1016/0277-9536\(95\)00116-O](http://dx.doi.org/10.1016/0277-9536(95)00116-O)
274. Moore JT, Silimperi DR, Bobula JA. Recognition of depression by family medicine residents: the impact of screening. *J Fam Pract* 1978;**7**:509–13.
275. Hoepfer EW, Nycz GR, Kessler LG, Burke JD, Pierce WE. The usefulness of screening for mental illness. *Lancet* 1984;**1**:33–5. [http://dx.doi.org/10.1016/S0140-6736\(84\)90192-2](http://dx.doi.org/10.1016/S0140-6736(84)90192-2)
276. German PS, Shapiro S, Skinner EA, Von Korff M, Klein LE, Turner RW, *et al.* Detection and management of mental health problems of older patients by primary care providers. *JAMA* 1987;**257**:489–93. <http://dx.doi.org/10.1001/jama.1987.03390040105026>
277. Rubenstein LV, Calkins DR, Young RT, Cleary PD, Fink A, Kosecoff J, *et al.* Improving patient function: a randomized trial of functional disability screening. *Ann Intern Med* 1989;**111**:836–42. <http://dx.doi.org/10.7326/0003-4819-111-10-836>
278. Thombs BD, Ziegelstein RC. Does depression screening improve depression outcomes in primary care? *BMJ* 2014;**348**:g1253. <http://dx.doi.org/10.1136/bmj.g1253>
279. Lambert MJ, Hansen NB, Finch AE. Patient-focused research: using patient outcome data to enhance treatment effects. *J Consult Clin Psychol* 2001;**69**:159–72. <http://dx.doi.org/10.1037/0022-006X.69.2.159>
280. Sapyta J, Riemer M, Bickman L. Feedback to clinicians: theory, research, and practice. *J Clin Psychol* 2005;**61**:145–53. <http://dx.doi.org/10.1002/jclp.20107>
281. Bickman L, Riemer M, Breda C, Kelley S. CFIT: a system to provide a continuous quality improvement infrastructure through organizational responsiveness, measurement, training, and feedback. *Rep Emotional Behav Disord Youth* 2006;**6**:86–9.

282. Wagner AK, Ehrenberg BL, Tran TA, Bungay KM, Cynn DJ, Rogers WH. Patient-based health status measurement in clinical practice: a study of its impact on epilepsy patients' care. *Qual Life Res* 1997;**6**:329–41. <http://dx.doi.org/10.1023/A:1018479209369>
283. Søreide K, Søreide AH. Using patient-reported outcome measures for improved decision-making in patients with gastrointestinal cancer – the last clinical frontier in surgical oncology? *Front Oncol* 2013;**3**:157. <http://dx.doi.org/10.3389/fonc.2013.00157>
284. Blakeman T, Chew-Graham C, Reeves D, Rogers A, Bower P. The Quality and Outcomes Framework and self-management dialogue in primary care consultations: a qualitative study. *Br J Gen Pract* 2011;**61**:e666–73. <http://dx.doi.org/10.3399/bjgp11X601389>
285. Mitcheson J, Cowley S. Empowerment or control? An analysis of the extent to which client participation is enabled during health visitor/client interactions using a structured health needs assessment tool. *Int J Nurs Stud* 2003;**40**:413–26. [http://dx.doi.org/10.1016/S0020-7489\(02\)00107-4](http://dx.doi.org/10.1016/S0020-7489(02)00107-4)
286. Coulter A, Roberts S, Dixon A. *Delivering Better Services for People with Long Term Conditions: Building the House of Care*. London: The King's Fund; 2013.
287. Santana MJ, Feeny D. Framework to assess the effects of using patient-reported outcome measures in chronic care management. *Qual Life Res* 2014;**23**:1505–13. <http://dx.doi.org/10.1007/s11136-013-0596-1>
288. Feldman-Stewart D, Brundage MD. A conceptual framework for patient-provider communication: a tool in the PRO research tool box. *Qual Life Res* 2009;**18**:109–14. <http://dx.doi.org/10.1007/s11136-008-9417-3>
289. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. *Patient Educ Couns* 2009;**74**:295–301. <http://dx.doi.org/10.1016/j.pec.2008.11.015>
290. Ziegler L, Harley C, Holch P, Keding A, Bamforth L, Warrington L, et al. Towards safer delivery and monitoring of cancer treatments. Electronic patient self-Reporting of Adverse-events: Patient Information and aDvice (eRAPID). *Psychooncology* 2012;**21**:15.
291. Warrington L, Absolom K, Velikova G. Integrated care pathways for cancer survivors – a role for patient-reported outcome measures and health informatics. *Acta Oncol* 2015;**54**:600–8. <http://dx.doi.org/10.3109/0284186X.2014.995778>
292. Chang CH, Cella D, Masters GA, Laliberte N, O'Brien P, Peterman A, Shervin D. Real-time clinical application of quality-of-life assessment in advanced lung cancer. *Clin Lung Cancer* 2002;**4**:104–9. <http://dx.doi.org/10.3816/CLC.2002.n.020>
293. Jensen RE, Snyder CF, Abernethy AP, Basch E, Potosky AL, Roberts AC, et al. Review of electronic patient-reported outcomes systems used in cancer clinical care. *J Oncol Pract* 2014;**10**:e215–22. <http://dx.doi.org/10.1200/JOP.2013.001067>
294. Snyder CF, Jensen R, Courtin SO, Wu AW. Website for Outpatient QOL Assessment Research Network. PatientViewpoint: a website for patient-reported outcomes assessment. *Qual Life Res* 2009;**18**:793–800. <http://dx.doi.org/10.1007/s11136-009-9497-8>
295. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. *Med Care* 2013;**51**(Suppl. 8):73–9. <http://dx.doi.org/10.1097/MLR.0b013e31829b1d84>
296. Thombs BD, Arthurs E, El-Baalbaki G, Meijer A, Ziegelstein RC, Steele RJ. Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: systematic review. *BMJ* 2011;**343**:d4825. <http://dx.doi.org/10.1136/bmj.d4825>

297. Thombs BD, Coyne JC, Cuijpers P, de Jonge P, Gilbody S, Ioannidis JP, et al. Rethinking recommendations for screening for depression in primary care. *CMAJ* 2012;**184**:413–18. <http://dx.doi.org/10.1503/cmaj.111035>
298. Benedetti F, Lanotte M, Lopiano L, Colloca L. When words are painful: unraveling the mechanisms of the placebo effect. *Neuroscience* 2007;**147**:260–71. <http://dx.doi.org/10.1016/j.neuroscience.2007.02.020>
299. Goldberg D. Use of the general health questionnaire in clinical work. *BMJ* 1986;**293**:1188–9. <http://dx.doi.org/10.1136/bmj.293.6556.1188>
300. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;**4**:293–307. <http://dx.doi.org/10.1007/BF01593882>
301. Hahn EA, Cella D, Chassany O, Fairclough DL, Wong GY, Hays RD, Clinical Significance Consensus Meeting Group. Precision of health-related quality-of-life data compared with other clinical measures. *Mayo Clin Proc* 2007;**82**:1244–54. <http://dx.doi.org/10.4065/82.10.1244>
302. Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. *Soc Sci Med* 1995;**40**:903–18. [http://dx.doi.org/10.1016/0277-9536\(94\)00155-M](http://dx.doi.org/10.1016/0277-9536(94)00155-M)
303. Lohr KN, Zebrack BJ. Using patient-reported outcomes in clinical practice: challenges and opportunities. *Qual Life Res* 2009;**18**:99–107. <http://dx.doi.org/10.1007/s11136-008-9413-7>
304. Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *JAMA* 1994;**272**:619–26. <http://dx.doi.org/10.1001/jama.1994.03520080061045>
305. Howell D, Mayo S, Currie S, Jones G, Boyle M, Hack T, et al. Psychosocial health care needs assessment of adult cancer patients: a consensus-based guideline. *Support Care Cancer* 2012;**20**:3343–54. <http://dx.doi.org/10.1007/s00520-012-1468-x>
306. Brow JP, McGee HM, O'Boyle CA. Conceptual approaches to the assessment of quality of life. *Psychol Health* 1997;**12**:737–51. <http://dx.doi.org/10.1080/08870449708406736>
307. Macduff C. Respondent-generated quality of life measures: useful tools for nursing or more fool's gold? *J Adv Nurs* 2000;**32**:375–82. <http://dx.doi.org/10.1046/j.1365-2648.2000.01486.x>
308. Pilnick A, Dingwall R. On the remarkable persistence of asymmetry in doctor/patient interaction: a critical review. *Soc Sci Med* 2011;**72**:1374–82. <http://dx.doi.org/10.1016/j.socscimed.2011.02.033>
309. Lupton D. The digitally engaged patient: self monitoring and self-care in the digital health era. *Soc Theory Health* 2013;**11**:256–70. <http://dx.doi.org/10.1057/sth.2013.10>
310. Lambert MJ, Harmon C, Slade K, Whipple JL, Hawkins EJ. Providing feedback to psychotherapists on their patients' progress: clinical results and practice suggestions. *J Clin Psychol* 2005;**61**:165–74. <http://dx.doi.org/10.1002/jclp.20113>
311. Greenhalgh J, Flynn R, Long AF, Tyson S. Tacit and encoded knowledge in the use of standardised outcome measures in multidisciplinary team decision making: a case study of in-patient neurorehabilitation. *Soc Sci Med* 2008;**67**:183–94. <http://dx.doi.org/10.1016/j.socscimed.2008.03.006>
312. Tyson SF, Greenhalgh J, Long AF, Flynn R. The influence of objective measurement tools on communication and clinical decision making in neurological rehabilitation. *J Eval Clin Pract* 2012;**18**:216–24. <http://dx.doi.org/10.1111/j.1365-2753.2010.01555.x>
313. Aktas A, Hullihen B, Shrotriya S, Thomas S, Walsh D, Estfan B. Connected health: cancer symptom and quality-of-life assessment using a tablet computer: a pilot study. *Am J Hosp Palliat Care* 2015;**32**:189–97. <http://dx.doi.org/10.1177/1049909113510963>

314. Holch P, Warrington L, Potrata B, Ziegler L, Absolom K, Hector C, *et al.* Acceptability of an online system for reporting and managing symptomatic adverse events (eRAPID): patients views. *Psychooncology* 2013;**22**:9–10.
315. Farmer A, Toms C, Hardinge M, Williams V, Rutter H, Tarassenko L. Self-management support using an Internet-linked tablet computer (the EDGE platform)-based intervention in chronic obstructive pulmonary disease: protocol for the EDGE-COPD randomised controlled trial. *BMJ Open* 2014;**4**:e004437. <http://dx.doi.org/10.1136/bmjopen-2013-004437>
316. Berry DL, Hong F, Halpenny B, Partridge A, Fox E, Fann JR, *et al.* The electronic self report assessment and intervention for cancer: promoting patient verbal reporting of symptom and quality of life issues in a randomized controlled trial. *BMC Cancer* 2014;**14**:513. <http://dx.doi.org/10.1186/1471-2407-14-513>
317. Berry DL, Hong F, Halpenny B, Partridge AH, Fann JR, Wolpin S, *et al.* Electronic self-report assessment for cancer and self-care support: results of a multicenter randomized trial. *J Clin Oncol* 2014;**32**:199–205. <http://dx.doi.org/10.1200/JCO.2013.48.6662>
318. Cartwright M, Hirani SP, Rixon L, Beynon M, Doll H, Bower P, *et al.* Effect of telehealth on quality of life and psychological outcomes over 12 months (Whole Systems Demonstrator telehealth questionnaire study): nested study of patient reported outcomes in a pragmatic, cluster randomised controlled trial. *BMJ* 2013;**346**:f653. <http://dx.doi.org/10.1136/bmj.f653>
319. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. *J Am Med Inform Assoc* 2007;**14**:269–77. <http://dx.doi.org/10.1197/jamia.M2270>
320. Glynn LG, Murphy AW, Smith SM, Schroeder K, Fahey T. Self-monitoring and other non-pharmacological interventions to improve the management of hypertension in primary care: a systematic review. *Br J Gen Pract* 2010;**60**:e476–88. <http://dx.doi.org/10.3399/bjgp10X544113>
321. Heritage J, Maynard DW. Introduction: Analyzing Interaction between Doctors and Patients in Primary Care Encounters. In Heritage J, Maynard DW, editors. *Communication in Medical Care: Interaction between Primary Care Physicians and Patients*. Cambridge: Cambridge University Press; 2006. pp. 1–21. <http://dx.doi.org/10.1017/CBO9780511607172.003>
322. Trujols J, Portella MJ, Iraurgi I, Campins MJ, Siñol N, de Los Cobos JP. Patient-reported outcome measures: are they patient-generated, patient-centred or patient-valued? *J Ment Health* 2013;**22**:555–62. <http://dx.doi.org/10.3109/09638237.2012.734653>
323. Jocham HR, Dassen T, Widdershoven G, Halfens R. Quality of life in palliative care cancer patients: a literature review. *J Clin Nurs* 2006;**15**:1188–95. <http://dx.doi.org/10.1111/j.1365-2702.2006.01274.x>
324. Bausewein C, Daveson BA, Currow DC, Downing J, Deliens L, Radbruch L, *et al.* EAPC White Paper on outcome measurement in palliative care: improving practice, attaining outcomes and delivering quality services – recommendations from the European Association for Palliative Care (EAPC) Task Force on Outcome Measurement. *Palliat Med* 2016;**30**:6–22. <http://dx.doi.org/10.1177/0269216315589898>
325. Snyder CF, Herman JM, White SM, Luber BS, Blackford AL, Carducci MA, Wu AW. When using patient-reported outcomes in clinical practice, the measure matters: a randomized controlled trial. *J Oncol Pract* 2014;**10**:e299–306. <http://dx.doi.org/10.1200/JOP.2014.001413>
326. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *JAMA* 2002;**288**:3027–34. <http://dx.doi.org/10.1001/jama.288.23.3027>

327. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, Selby PJ. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol* 2004;**22**:714–24. <http://dx.doi.org/10.1200/JCO.2004.06.078>
328. Nilsson E, Wenemark M, Bendtsen P, Kristenson M. Respondent satisfaction regarding SF-36 and EQ-5D, and patients' perspectives concerning health outcome assessment within routine health care. *Qual Life Res* 2007;**16**:1647–54. <http://dx.doi.org/10.1007/s11136-007-9263-8>
329. Neudert C, Wasner M, Borasio GD. Patients' assessment of quality of life instruments: a randomised study of SIP, SF-36 and SEIQoL-DW in patients with amyotrophic lateral sclerosis. *J Neurol Sci* 2001;**191**:103–9. [http://dx.doi.org/10.1016/S0022-510X\(01\)00612-8](http://dx.doi.org/10.1016/S0022-510X(01)00612-8)
330. Hickey AM, Bury G, O'Boyle CA, Bradley F, O'Kelly FD, Shannon W. A new short form individual quality of life measure (SEIQoL-DW): application in a cohort of individuals with HIV/AIDS. *BMJ* 1996;**313**:29–33. <http://dx.doi.org/10.1136/bmj.313.7048.29>
331. Mallinson S. Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire. *Soc Sci Med* 2002;**54**:11–21. [http://dx.doi.org/10.1016/S0277-9536\(01\)00003-X](http://dx.doi.org/10.1016/S0277-9536(01)00003-X)
332. Westerman M, Hak T, The AM, Groen H, van der Wal G. Problems eliciting cues in SEIQoL-DW: quality of life areas in small-cell lung cancer patients. *Qual Life Res* 2006;**15**:441–9. <http://dx.doi.org/10.1007/s11136-005-2831-x>
333. O'Boyle CA, Browne J, Hickey A, McGee H, Joyce CRB. *Schedule for the Evaluation of Individual Quality of Life (SEIQoL): a Direct Weighting Procedure for Quality Of Life Domains (SEQoL-DW). Administration Manual*. Dublin: Royal College of Surgeons of Ireland; 1995.
334. Farquhar M, Ewing G, Higginson IJ, Booth S. The experience of using the SEIQoL-DW with patients with advanced chronic obstructive pulmonary disease (COPD): issues of process and outcome. *Qual Life Res* 2010;**19**:619–29. <http://dx.doi.org/10.1007/s11136-010-9631-7>
335. Snyder CF, Jensen RE, Geller G, Carducci MA, Wu AW. Relevant content for a patient-reported outcomes questionnaire for use in oncology clinical practice: putting doctors and patients on the same page. *Qual Life Res* 2010;**19**:1045–55. <http://dx.doi.org/10.1007/s11136-010-9655-z>
336. Leydon GM, Dowrick CF, McBride AS, Burgess HJ, Howe AC, Clarke PD, *et al*. Questionnaire severity measures for depression: a threat to the doctor-patient relationship? *Br J Gen Pract* 2011;**61**:117–23. <http://dx.doi.org/10.3399/bjgp11X556236>
337. Pettersson A, Björkelund C, Petersson EL. To score or not to score: a qualitative study on GPs views on the use of instruments for depression. *Fam Pract* 2014;**31**:215–21. <http://dx.doi.org/10.1093/fampra/cmt082>
338. Cheyne A, Kinn S. Counsellors' perspectives on the use of the Schedule for the Evaluation of Individual Quality of Life (SEIQoL) in an alcohol counselling setting. *Brit J Guid Couns* 2001;**29**:35–46. <http://dx.doi.org/10.1080/03069880020019383>
339. Hall CL, Taylor J, Moldavsky M, Marriott M, Pass S, Newell K, *et al*. A qualitative process evaluation of electronic session-by-session outcome measurement in child and adolescent mental health services. *BMC Psychiatry* 2014;**14**:113. <http://dx.doi.org/10.1186/1471-244X-14-113>
340. Wolpert M, Curtis-Tyler K, Edbrooke-Childs J. A qualitative exploration of patient and clinician views on patient reported outcome measures in child mental health and diabetes services. *Adm Policy Ment Health* 2016;**43**:309–15. <http://dx.doi.org/10.1007/s10488-014-0586-9>
341. Stasiak K, Parkin A, Seymour F, Lambie I, Crengle S, Pasene-Mizziebo E, Merry S. Measuring outcome in child and adolescent mental health services: consumers' views of measures. *Clin Child Psychol Psychiatry* 2013;**18**:519–35. <http://dx.doi.org/10.1177/1359104512460860>

342. Hagelin CL, Wengström Y, Tishelman C, Fürst CJ. Nurses' experiences of clinical use of a quality of life instrument in palliative care. *Contemp Nurse* 2007;**27**:29–38. <http://dx.doi.org/10.5172/conu.2007.27.1.29>
343. Slater A, Freeman E. Patients' views of using an outcome measure in palliative day care: a focus group study. *Int J Palliat Nurs* 2004;**10**:343–51. <http://dx.doi.org/10.12968/ijpn.2004.10.7.14576>
344. Hughes R, Aspinall F, Addington-Hall J, Chidgey J, Drescher U, Dunckley M, Higginson IJ. Professionals' views and experiences of using outcome measures in palliative care. *Int J Palliat Nurs* 2003;**9**:234–8. <http://dx.doi.org/10.12968/ijpn.2003.9.6.11511>
345. McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Med* 2011;**9**:86. <http://dx.doi.org/10.1186/1741-7015-9-86>
346. Hughes R, Aspinall F, Addington-Hall JM, Dunckley M, Faull C, Higginson I. It just didn't work: the realities of quality assessment in the English health care context. *Int J Nurs Stud* 2004;**41**:705–12. <http://dx.doi.org/10.1016/j.ijnurstu.2004.02.005>
347. Eischens MJ, Elliott BA, Elliott TE. Two hospice quality of life surveys: a comparison. *Am J Hosp Palliat Care* 1998;**15**:143–8. <http://dx.doi.org/10.1177/104990919801500308>
348. Gamlen E, Arber A. First assessments by specialist cancer nurses in the community: an ethnography. *Eur J Oncol Nurs* 2013;**17**:797–801. <http://dx.doi.org/10.1016/j.ejon.2013.03.006>
349. Bensing J. Bridging the gap. The separate worlds of evidence-based medicine and patient-centered medicine. *Patient Educ Couns* 2000;**39**:17–25. [http://dx.doi.org/10.1016/S0738-3991\(99\)00087-7](http://dx.doi.org/10.1016/S0738-3991(99)00087-7)
350. Fagerlind H, Lindblad AK, Bergström I, Nilsson M, Nauclet G, Glimelius B, Ring L. Patient-physician communication during oncology consultations. *Psychooncology* 2008;**17**:975–85. <http://dx.doi.org/10.1002/pon.1410>
351. Fayers PM. Evaluating the effectiveness of using PROs in clinical practice: a role for cluster-randomised trials. *Qual Life Res* 2008;**17**:1315–21. <http://dx.doi.org/10.1007/s11136-008-9391-9>
352. Velikova G, Keding A, Harley C, Cocks K, Booth L, Smith AB, et al. Patients report improvements in continuity of care when quality of life assessments are used routinely in oncology practice: secondary outcomes of a randomised controlled trial. *Eur J Cancer* 2010;**46**:2381–8. <http://dx.doi.org/10.1016/j.ejca.2010.04.030>
353. Mills ME, Murray LJ, Johnston BT, Cardwell C, Donnelly M. Does a patient-held quality-of-life diary benefit patients with inoperable lung cancer? *J Clin Oncol* 2009;**27**:70–7. <http://dx.doi.org/10.1200/JCO.2008.17.5687>
354. Berry DL, Blumenstein BA, Halpenny B, Wolpin S, Fann JR, Austin-Seymour M, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. *J Clin Oncol* 2011;**29**:1029–35. <http://dx.doi.org/10.1200/JCO.2010.30.3909>
355. Detmar SB, Aaronson NK, Wever LD, Muller M, Schornagel JH. How are you feeling? Who wants to know? Patients' and oncologists' preferences for discussing health-related quality-of-life issues. *J Clin Oncol* 2000;**18**:3295–301.
356. Taylor S, Harley C, Campbell LJ, Bingham L, Podmore EJ, Newsham AC, et al. Discussion of emotional and social impact of cancer during outpatient oncology consultations. *Psychooncology* 2011;**20**:242–51. <http://dx.doi.org/10.1002/pon.1730>
357. Velikova G, Awad N, Coles-Gale R, Wright EP, Brown JM, Selby PJ. The clinical value of quality of life assessment in oncology practice – a qualitative study of patient and physician views. *Psychooncology* 2008;**17**:690–8. <http://dx.doi.org/10.1002/pon.1295>

358. Absolom K, Holch P, Pini S, Hill K, Liu A, Sharpe M, *et al.* The detection and management of emotional distress in cancer patients: the views of health-care professionals. *Psychooncology* 2011;**20**:601–8. <http://dx.doi.org/10.1002/pon.1916>
359. Rogers MS, Todd CJ. The ‘right kind’ of pain: talking about symptoms in outpatient oncology consultations. *Palliat Med* 2000;**14**:299–307. <http://dx.doi.org/10.1191/026921600669288537>
360. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Role of health-related quality of life in palliative chemotherapy treatment decisions. *J Clin Oncol* 2002;**20**:1056–62. <http://dx.doi.org/10.1200/JCO.20.4.1056>
361. Henly SJ. The promise of PROMIS. *Nurs Res* 2010;**59**:77. <http://dx.doi.org/10.1097/NNR.0b013e3181d7d1b1>
362. Wong CK, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. *Realist Synthesis: Quality Standards for Researchers and Peer Reviewers*. 2014. URL: www.ramesesproject.org/index.php?pr=Project_outputs (accessed 23 September 2015).
363. Ashworth M, White P, Jongsma H, Schofield P, Armstrong D. Antibiotic prescribing and patient satisfaction in primary care in England: cross-sectional analysis of national patient survey and prescribing data. *Br J Gen Pract* 2016;**66**:e40–6. <http://dx.doi.org/10.3399/bjgp15x688105>
364. Basser MR. *PROMS Benefits Case Study: Patient Reported Outcome Measure (PROMs) Outputs*. London: HSCIC; 2015.
365. Keetharuth A, Mulhern B, Wong R, Ara R, Franklin M, Jones G, *et al.* *Supporting the Routine Collection of Patient Reported Outcome Measures (PROMs) for the National Clinical Audit. Work Package 2. How Should PROMs Data be Collected? Report to the Department of Health Policy Research Programme*. Sheffield: University of Sheffield and University of York; 2015.
366. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* 2004;**82**:581–629. <http://dx.doi.org/10.1111/j.0887-378X.2004.00325.x>
367. Best A, Greenhalgh T, Lewis S, Saul JE, Carroll S, Bitz J. Large-system transformation in health care: a realist review. *Milbank Q* 2012;**90**:421–56. <http://dx.doi.org/10.1111/j.1468-0009.2012.00670.x>
368. Blakeman T, Bower P, Reeves D, Chew-Graham C. Bringing self-management into clinical view: a qualitative study of long-term condition management in primary care consultations. *Chronic Illn* 2010;**6**:136–50. <http://dx.doi.org/10.1177/1742395309358333>
369. Rhodes P, Langdon M, Rowley E, Wright J, Small N. What does the use of a computerized checklist mean for patient-centered care? The example of a routine diabetes review. *Qual Health Res* 2006;**16**:353–76. <http://dx.doi.org/10.1177/1049732305282396>
370. Hall CL, Moldavsky M, Taylor J, Sayal K, Marriott M, Batty MJ, *et al.* Implementation of routine outcome measurement in child and adolescent mental health services in the United Kingdom: a critical perspective. *Eur Child Adolesc Psychiatry* 2014;**23**:239–42. <http://dx.doi.org/10.1007/s00787-013-0454-2>
371. Palmer D, El Miedany Y. PROMs: a novel approach to arthritis self-management. *Br J Nurs* 2012;**21**:601–2, 605–7. <http://dx.doi.org/10.12968/bjon.2012.21.10.601>
372. Protheroe J, Blakeman T, Bower P, Chew-Graham C, Kennedy A. An intervention to promote patient participation and self-management in long term conditions: development and feasibility testing. *BMC Health Serv Res* 2010;**10**:206. <http://dx.doi.org/10.1186/1472-6963-10-206>

373. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. *J Med Internet Res* 2014;**16**:e295. <http://dx.doi.org/10.2196/jmir.3348>
374. Dorsch MP, Farris KB, Bleske BE, Koelling TM. A web application for self-monitoring improves symptoms in chronic systolic heart failure. *Telemed J E Health* 2015;**21**:267–70. <http://dx.doi.org/10.1089/tmj.2014.0095>
375. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, *et al*. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *BMJ* 2012;**344**:e3874. <http://dx.doi.org/10.1136/bmj.e3874>
376. Longo DR, Land G, Schramm W, Fraas J, Hoskins B, Howell V. Consumer reports in health care. Do they make a difference in patient care? *JAMA* 1997;**278**:1579–84. <http://dx.doi.org/10.1001/jama.1997.03550190043042>

Appendix 1 Search strategies

One example strategy for each electronic subject-based search is presented. The following strategies were closely translated for use in other databases listed in the main report.

Patient-reported outcome measures feedback reviews

Ovid MEDLINE(R)

Date range searched: 1946 to week 3, April 2014.

Date of search: 2 May 2014.

Search strategy

1. ("patient report*" adj3 (outcome* or data)).tw. (3342)
2. proms.tw. (151)
3. ("public* report*" adj3 (outcome* or data or information or care)).tw. (295)
4. exp *Health Status Indicators/ and (*patient satisfaction/ or *patient preference/ or *Patient-Centered Care/) (189)
5. *"Outcome Assessment (Health Care)"/ and (*patient satisfaction/ or *patient preference/ or *Patient-Centered Care/ or *Professional-Patient Relations/) (791)
6. *Self Report/ and *Quality of Life/ (101)
7. *"Quality of Life"/ and Health Status/ and patient satisfaction/ and ("Outcome Assessment (Health Care)"/ or Questionnaires/) (212)
8. ("public report*" adj3 hospital*).tw. (78)
9. (star adj2 rating*).ti. (22)
10. or/1-9 [PROMs] (4813)
11. exp Decision Making/ (122,197)
12. quality indicators, health care/ (9977)
13. exp treatment outcome/ (640,336)
14. exp Feedback/ (42,542)
15. ((PROM or PROMs or PRO) adj3 acceptab*).tw. (16)
16. (improve* or feedback).ti. (141,142)
17. or/11-16 [Feedback theories] (925,409)
18. 10 and 17 [Proms Feedback] (1865)
19. limit 18 to "reviews (maximizes specificity)" (143)
20. (Review* adj5 outcome*).ab. (13,974)
21. (Review* adj7 outcome*).ti. (2133)
22. (Review* adj2 (PRO or PROM or PROMs)).tw. (117)
23. (review* adj3 data).tw. (31,333)
24. or/20-23 (46,469)
25. 18 and 24 (81)
26. 19 or 25 (190)

Patient-reported outcome measures feedback opinions and commentaries

Ovid MEDLINE(R)

Date range searched: 1946 to week 3, April 2014.

Date of search: 2 May 2014.

Search strategy

1. Comment/ (535,128)
2. Letter/ (808,856)
3. Editorial/ (336,724)
4. news/ or newspaper article/ (169,377)
5. "Comment on".ti. (12,027)
6. (letter* adj3 editor*).ti. (3046)
7. opinion*.ti. (9839)
8. (view or views).ti. (37,996)
9. comment.cm. (1,007,256)
10. or/1-9 [Commentary and Opinion Pieces] (1,886,406)
11. ("patient report*" adj3 (outcome* or data)).tw. (3342)
12. proms.tw. (151)
13. ("public* report*" adj3 (outcome* or data or information or care)).tw. (295)
14. *"Outcome Assessment (Health Care)"/ and (*patient satisfaction/ or *patient preference/ or *Patient-Centered Care/ or *Professional-Patient Relations/) (791)
15. ("public report*" adj3 hospital*).tw. (78)
16. (star adj2 rating*).ti. (22)
17. or/11-16 [PROMs] (4403)
18. Patient Participation/ (17,911)
19. quality indicators, health care/ (9977)
20. exp treatment outcome/ (640,336)
21. (improve* or impact* or transform* or feedback).ti. (321,717)
22. exp Feedback/ (42,542)
23. *Attitude of health personnel/ (45,015)
24. Quality of healthcare/ (56,546)
25. or/18-24 [Feedback theories] (1,084,628)
26. 10 and 17 and 25 [Proms Feedback Commentaries] (286)

Provider feedback outcomes of patient-reported outcome measures or similar performance indicators

Ovid MEDLINE(R)

Date range searched: 1946 to week 2, October 2014.

Date of search: 26 September 2014.

Search strategy

1. ("patient report*" adj3 (outcome* or data)).tw. (4081)
2. proms.tw. (184)
3. exp *Health Status Indicators/ and (*patient satisfaction/ or *patient preference/ or *Patient-Centered Care/) (201)
4. *"Outcome Assessment (Health Care)"/ and (*patient satisfaction/ or *patient preference/ or *Patient-Centered Care/ or *Professional-Patient Relations/) (868)
5. *Self Report/ and *Quality of Life/ (147)
6. *"Quality of Life"/ and Health Status/ and patient satisfaction/ and ("Outcome Assessment (Health Care)"/ or Questionnaires/) (227)
7. (star adj2 rating*).ti. (23)
8. "quality and outcomes framework".tw. (276)
9. *"Outcome Assessment (Health Care)"/mt, og (5763)
10. *Questionnaires/ and *Quality of Life/ (4913)

11. *quality indicators, health care/ (6099)
12. (performance adj5 feedback).tw. (2343)
13. (performance adj3 results).tw. (7704)
14. (performance adj3 indicator*).tw. (2617)
15. ("hospital performance" adj2 data).tw. (43)
16. benchmark*.tw. (15,827)
17. Benchmarking/ (10,541)
18. (dashboard* or "dash board*").tw. (339)
19. (audit or audited or auditing).tw. (24,386)
20. *Management Audit/ or exp *Clinical Audit/ (9076)
21. *Total Quality Management/mt (1260)
22. (public* report* adj3 (performance or quality)).tw. (292)
23. or/1-22 [PROMs and other quality indicators] (83,107)
24. exp Decision Making/ (130,090)
25. Decision*.ti. (39,917)
26. Feedback/ (27,366)
27. (improve* or feedback).ti. (151,427)
28. ((public or publically) adj1 report*).tw. (908)
29. "Attitude of Health Personnel"/ (94,489)
30. (improv* adj2 participat*).tw. (941)
31. communication/ (64,920)
32. or/24-31 [Agg Outcomes of Feedback] (460,306)
33. 23 and 32 [Proms Feedback] (9429)
34. qualitative*.tw. (143,688)
35. qualitative research/ (21,975)
36. survey*.ti. (94,110)
37. attitude*.ti. (33,395)
38. (barrier* or faciliator*).ti. (31,183)
39. limit 33 to "qualitative (maximizes specificity)" (495)
40. or/34-38 [Qualitative] (303,577)
41. 23 and 32 and 40 (824)
42. 39 or 41 (895)

Search MEDLINE and EMBASE search for patient experience data (example from MEDLINE given)

Database: Ovid MEDLINE(R)

Date range searched: 1946 to week 2, November 2014.

Date of search: 20 November 2014.

Search strategy

1. inpatient survey.tw. (50)
2. inpatient experience survey.tw. (2)
3. GP survey.tw. (25)
4. GP experience survey.tw. (0)
5. (inpatient adj3 survey).tw. (157)
6. (GP adj3 survey).tw. (125)
7. 5 or 6 (281)
8. experience.tw. (429,041)
9. 7 and 8 (36)
10. 1 or 2 or 3 or 4 or 9 (102)

Appendix 2 Example of a data extraction form

Pham *et al.*²²⁸

Aims

Explore the impact of different quality reporting programmes on hospital operations, focusing on the care of patients with congestive heart failure.

Indicator

This paper explores provider responses to a range of performance reporting initiatives, but pays particular attention to two initiatives. The first is the reporting required by the JCAHO, an organisation responsible for the accreditation and regulation of hospitals in the USA; public reporting is a condition of accreditation. By contrast, the Hospital Quality Initiative is a voluntary system launched by the Centres for Medicare and Medicaid Services and is reported on the Hospital Compare website. The authors report that when it was initially launched, participation was poor until the Medicare Prescription Drug, Improvement and Modernization Act established that non-participating hospitals would not receive a 0.4% annual payment update and now almost all eligible hospitals are participating.

This illustrates the role of incentives in prompting hospitals to participate in reporting schemes and also their possible role in how hospitals respond to this reporting. In analysing their findings, the authors identified a number of axes along which different reporting programmes vary, which correspond quite closely to the contextual factors identified within the public disclosure, audit and feedback and benchmarking theories. These included (1) sponsorship – by purchaser, regulator, private insurer, professional groups or other private organisations; (2) data type: hospitals submit primary data for public reporting (JCAHO), primary data for private benchmarking (e.g. hospital consortia); (3) secondary data (e.g. claims or patient surveys); (4) mandatory versus voluntary – although most programmes are voluntary, this is influenced by the nature of incentives attached to the programme – incentives explicitly tied to participation in the programme perceived as rewards (e.g. pay for performance), or punitive (e.g. loss of accreditation); (5) QI support – whether or not programmes provide prescriptive information to guide hospitals QI activities; and (6) inclusion of clinical outcome measures.

Methodology

The paper draws on data collected as part of the Community Tracking Study. The Community Tracking Study is a longitudinal study based on site visits and surveys of health-care purchasers, insurers and providers and is focused on tracking changes in the accessibility, cost and quality of health care. The data selected for the study were collected in 2004–5 and consisted of 111 interviews with five hospital association leaders, representatives from the JCAHO and CMS and six state reporting programmes, 21 chief executive officers, 21 vice presidents of nursing, 30 quality officers, 26 clinical directors from 2–4 of the largest hospitals or hospital systems in 12 health-care markets across the USA. The data were collected using semistructured interviews that explored specific reporting programmes and their perceived impact on the hospital's organisational culture around QI, priorities, budget, data collection and review activities, feedback and accountability mechanisms. Clinical directors were also asked about their use of 11 QI tools targeted at chronic heart failure, as the CMS and JCAHO reporting systems both included chronic heart failure.

Findings

Hospitals were currently responding to multiple reporting programmes; all were involved in CMS and JCAHO reporting, but also indicated that they were involved in additional programmes, with hospitals involved in up to seven additional programmes. In all, respondents mentioned 38 programmes, which varied according to sponsorship, type of data collected and the degree of support offered by the programme.

Senior hospital leaders perceived that key drivers such as linkages to payment, JCAHO accreditation and peer pressure from public benchmarking have raised the priority given to quality measurement and improvement by hospital leaders. This was manifest in several ways including inclusion of QI priorities in strategic planning, boards and senior management taking more responsibility for the formal review of performance data and associated QI strategies, and performance-related pay for chief executives.

Respondents also believed that CMS and JCAHO reporting programmes had positively changed doctors' attitudes towards quality measurement and improvement. The accreditation and financial consequences of reporting programmes could be used as leverage by quality officers in their dealings with physicians. Quality officers believed that doctors liked having a set of clinical priorities to guide their work and that doctors responded to performance feedback due to peer pressure from public reporting of benchmarked data, supporting benchmarking theories.

Hospitals that had a longstanding involvement in QI activities did not feel that the requirement of the JCAHO and CMS had a significant impact on their QI priorities (links to Smith *et al.*²³⁰). Senior leaders in most hospitals felt that the reporting requirements by the CMS and JCAHO centred on a limited number of priorities, which focused attention away from other important areas (similar to tunnel vision reported in other studies).

As a result of their mandatory requirements, participants felt that the drive to participate in CMS and JCAHO reporting was of a 'push' nature. Hospitals did direct resources at the reported clinical areas but 'without taking standardized approaches to improving performance'. Motivation to be involved in other reporting initiatives was described more in terms of a 'pull' because they offered support by specifying changes in care processes. These forms of reporting were seen as attractive because they 'don't leave hospitals flailing about trying to identify evidence-based interventions on their own and because they encourage a culture of continuous QI (similar to Smith *et al.*²³⁰). These sorts of programmes were mostly likely to be those QI organisations, state QI organisations and professional organisations.

Hospital systems sent individual faculties their disaggregated performance scores and most hospitals reported that they share their overall performance scores with the entire staff.

Hospital respondents were divided on whether or not reporting had a significant impact on specific process changes to improve the quality of care. Those who had participated in QI programmes prior to being involved in public report programmes felt that public reporting had little impact on their QI interventions as these were largely driven internally or by both reporting and internally. Respondents were also divided on whether reporting had a 'spillover' effect on improving quality for the non-reporting conditions; some believed QI was limited to the targeted conditions, while others felt that it had raised frontline staff's ability and eagerness to identify and address problems in non-reported areas too.

The authors draw some interesting conclusions about the contextual factors that influence provider motivation to participate in and respond to public reporting. They argue that 'National programs with deep payment or regulatory leverage can influence nearly all hospitals and garner attention from those that would otherwise not prioritise QI highly . . . Voluntary programs, particularly those offering ongoing QI support, help focus priorities at hospitals that are eager to take on the more challenging goals of QI'.

Thus, national programmes with mandatory reporting and regulatory or payment consequences have increased hospital leadership as well as frontline staff's attention to quality and directed resources towards QI. However, at the same time, such public reporting initiatives may 'artificially narrow the scope of QI in which hospitals might otherwise engage, especially for those with long institutional histories of QI'.

Appendix 3 Example of one 'theory theme'

Formal incentives/sanctions and their impact on whether hospitals respond

Longo et al., 1997³⁷⁶

Hospitals in a competitive market were more likely to make changes to policies. Report cards backed by government legislation put providers in a position where they have to respond or face sanctions.

Mehrotra et al., 2003¹⁹⁵

Hospitals with greater experience of public reporting are more likely to be in favour of financial incentives for good performance in order to incentivise behaviour change and fund changes to practice – public disclosure theories.

The balance of power, determined by market conditions, between employers and providers determined the extent to which employers could force providers to participate in public reporting – public disclosure and benchmarking.

Mannion and Goddard, 2001²²⁰

In the two trusts that performed 'worse' on stroke indicators, they conducted an internal audit to check the quality of the data (i.e. to verify the indicators). In both 'worse' and 'average' trusts, CRAG data used alongside other data to argue for resources to support new services (e.g. stroke unit or breast clinics). Rarely served as a primary catalyst for QI. CRAG indicators were not part of formal incentive scheme and there was no pressure from external bodies to perform well/to held account for performance.

Pham et al., 2006²²⁸

Linkages to payment, JCAHO accreditation and peer pressure from public benchmarking raised the priority given to quality measurement and improvement by hospital leaders and frontline staff.

Authors conclude that 'National programs with deep payment or regulatory leverage can influence nearly all hospitals and garner attention from those that would otherwise not prioritise QI highly . . . Voluntary programs, particularly those offering ongoing QI support, help focus priorities at hospitals that are eager to take on the more challenging goals of QI'.

Wilkinson et al., 2000²¹⁹

Authors hypothesise that the absence of specific incentives to change (either for rewarding good performance or punishing poor performance) may have limited the extent to which GP practices responded to them.

Chassin et al., 2002¹⁹¹

Hospitals are unlikely to work together unless they are not in competition (e.g. they are geographically distant and not competing for the same patients). This relates to Wolfram Cox's model of benchmarking.

Geissler et al., 2013²³⁴

Incentives for improvement activities were the retention of patients or attracting new patients (this is clearly where there is a choice to be made). A barrier to completing the MHQP was the lack of (other) incentives.

Appendix 4 Table of theories tested and included studies

Programme theory	Linked subtheories	Number of studies
Chapter 4		
Theory 1: the feedback of PROMs or performance data leads to improve patient care	Theory 1a: PROMs data will lead to improved patient outcomes	2 ^{173,174}
	Theory 1b: the public reporting of performance data leads to improved patient outcomes	1 ²⁷
Theory 2: feedback influences providers' behaviour according to whether it is delivered privately (confidentially) or publicly	Theory 2a: providers respond to private/confidential feedback on performance	2 ^{176,177}
	Theory 2b: public reporting of performance places additional pressure on providers to respond	3 ^{35,39,118,179}
Theory 3: public reporting of poor performance threatens provider's market share and prompts a significant provider response	Theory 3a: providers experience a change in their market share following public report	4 ^{27,182,184,191}
	Theory 3b: providers take steps to improve the quality of care because they are worried about potential threats to their market share	
	Theory 3c: providers perceive that report cards pose a threat to their market share	3 ^{118,179,192,193}
Theory 4: providers perceived that report cards damage their professional or their hospitals' reputation	Theory 4a: providers perceive that report cards affect market share by damaging a hospitals reputation	4 ^{18,194–196}
	Theory 4b: media reports of hospital performance damage a hospital's reputation and affect their market share	4 ^{185,186,197,198}
Theory 5: providers respond through comparing themselves with their peers		3 ^{98,194,199}
Theory 6: consumers choose hospitals on the basis of information about the quality of services	Theory 6a: patients are willing to move hospitals in response to poor performance	6 ^{79,200,209–212}
	Theory 6b: patients choose hospitals on the basis of reports about service quality	
	Theory 6c: patients respond to poor hospital performance when it is report in the media	
Chapter 5		
Theory 7: financial incentives and sanctions influence providers' response to the public reporting of performance data	Theory 7a: financial incentives accelerate and amplify the impact of public reporting and feedback of performance data on improvements to patient care	4 ^{215–218}
	Theory 7b: providers do not make improvements to patient care in response to performance feedback when no financial incentives are attached	2 ^{219–221}
	Theory 7c: financial incentives attached to the feedback of performance data can lead to 'tunnel vision'	3 ^{18,80,81}
Theory 8: the perceived credibility of performance data influences providers' responses to the feedback of performance data	Theory 8a: the perceived credibility of the performance data influences providers' responses to performance data	3 ^{97,195,224}

Programme theory	Linked subtheories	Number of studies
Theory 9: the degree to which performance data are 'actionable' influences providers' responses to the feedback of performance data	Theory 8b: the source of performance data influences providers' perceptions of its credibility	3 ^{225–227}
	Theory 8c: the perceived underlying driver of public reporting systems influences providers' responses	4 ^{75,98,223,228}
	Theory 8d: clinicians have greater trust in clinician-led reporting systems	3 ^{199,229,230}
	Theory 9a: patient experience data are actionable and enable providers to take steps to improve patient care	7 ^{222,231–236}
	Theory 9b: making patient experience data more immediate and integrated into clinical discussion improves provider responses	2 ^{237,238}
	Theory 9c: providing tailored support to interpret and act on patient experience data improves providers' responses	
Chapter 8		
Theory 10: PROMs act as a tool to enable patients to raise or share issues with clinicians	Theory 10a: PROMs completion prompts patients to reflect on their own situation, which then enables them to identify and prioritise what is important to them and what they wish to share or raise with clinicians	23 ^{80,81,242–244,325,328,329,331,332,334,336–344,346–348}
	Theory 10b: PROMs completion reminds patients to mention or raise issues with clinicians	
	Theory 10c: PROMs completion gives patients permission to share or raise issues with clinicians	
Theory 11: PROMs completion may constrain the development of the patient–clinician relationship or hinder the flow of the consultation	Theory 11a: PROMs completion may constrain the development of the patient–clinician relationship	
	Theory 11b: PROMs completion or review may hinder the flow of the consultation	
Theory 12: the structure and format of the PROM make a difference to whether or not the intended outcomes of the PROM are achieved	Theory 12a: the length and format of the PROM makes a difference to whether or not the intended outcomes are achieved	
	Theory 12b: whether the PROM is individualised or standardised makes a difference to whether or not the intended outcomes are achieved	
Theory 13: whether the completion and review of PROMs supports or constrains the existing relationship depends on	Theory 13a: the existing relationship between patients and clinicians	
	Theory 13b: the point in the relationship-building process when the PROM is completed or reviewed	
	Theory 13c: patients' and clinicians' existing preferences for relationship building	
Theory 14: whether PROMs data collection is incentivised or used for other purposes makes a difference to the achievement of its intended outcomes		

Programme theory	Linked subtheories	Number of studies
Theory 15: PROMs feedback results in the increased discussion of HRQoL issues during the consultation and this promotes increased patient well-being		5 ^{63,326,327,352–354}
Theory 16: PROMs feedback increases the discussion of HRQoL issues during the consultation which in turn influences patient management and improves patients' well-being		
Theory 17: PROMs feedback does not change clinicians' communication practices during the consultation		
Theory 18: clinicians or patients do not consider it within their remit or judge that it is not appropriate to discuss psychosocial issues during the consultation		8 ^{65,244,355–360}
Theory 19: clinicians avoid discussion of issues that they feel they cannot or do not know how to treat		

Appendix 5 Dissemination plan

Please note that this plan is subject to change and revision.

1. A final report to the National Institute for Health Research.
2. A briefing document containing guidance to NHS managers summarising the findings and specifying how they might support the successful collation, interpretation and utilisation of PROMs data to improve the quality of patient care.
3. Sharing our report directly with the Organisation for Economic Co-operation and Development to inform their ongoing work on PROMs.
4. Showcasing our findings through local, national and international networks and conferences, including:
 - plenary presentation on project findings at the 1-day conference 'Advances in Patient Reported Outcomes Research and their application to decision making', hosted by the Yorkshire and Humber, Peninsula and East of England Collaborations for Leadership in Applied Health Research and Care, 9 June 2016, Sheffield, UK
 - poster presentation at the Health Services Research Network conference, 13–14 July 2016, Nottingham, UK
 - oral presentation at the 60th Annual Society for Social Medicine Conference, 14–16 September 2016, York, UK
 - plenary presentation at the 23rd International Society for Quality of Life Research (ISOQOL) annual conference, 19–22 October 2016, Copenhagen, Denmark.
5. Using social networks such as Twitter (Twitter Inc., San Francisco, CA, USA) to highlight our findings to a wide range of audiences.
6. A specific dissemination event to present our guidance to NHS decision-makers.
7. Peer-reviewed publications in academic journals and articles in practitioner and management journals and newsletters.
8. Working with PPI organisations to identify innovative ways of ensuring that our findings reach patients and the general public.

A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and flow.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library